Inclusive Intelligence: Artificial Intelligence in the Service of Science, Work, and the Public Good

Report, version 4: Revised October, 2019
Prepared for distribution to the UC Berkeley Campus Community

Executive Summary:
Artificial Intelligence and associated innovations in data science and machine learning are transforming how we work, learn, and communicate; how we govern; and how we live our daily lives. Although AI has great potential to streamline work and facilitate scientific discovery, it can also exacerbate socio-economic disparity, erode privacy, and concentrate power among a few corporations and countries. AI today is exclusionary in four ways:

1. Arcane: many AI techniques and results require advanced math and statistical knowledge at graduate-school levels.
2. Inaccessible: requiring access to substantial datasets and computation.
3. Centralizing: only a small number of companies and universities have the funding and expertise to advance AI.
4. Polarizing: increasing social disparities in knowledge, income, and privilege.

In response, UC Berkeley faculty are developing what we call **"Inclusive Intelligence",** a comprehensive research and teaching agenda for AI that is: 1) Inclusive of diverse disciplines and ways of knowing; 2) Inclusive of diverse populations including those most vulnerable; 3) Inclusive of the spectrum of human and artificial intelligences and the ways they can constructively interact, complement, and enhance each other.

UC Berkeley is uniquely positioned as a trusted global institution to lead this effort. Building on an unequaled legacy in research and pursuit of social justice and the public good, with an extremely popular and pioneering new undergraduate major in Data Science, Berkeley faculty, graduate students, and undergraduates will advance basic research and develop AI systems and policies to support effective, responsible applications to support science, work, and the public good.

Today, the field of AI is a patchwork of techniques lacking coherent, systematic design principles and tools. Analogous to the way that the application of genetic engineering evolved from the science of biology, what is needed is a systems perspective for AI that integrates basic science, data, and algorithms with human values. The Inclusive Intelligence research and education agenda will unite the diversity of academic strengths and the public mission of Berkeley to advance a vision of AI that centers humans in all their diversity, public values such

as social inclusion and human rights, and scientific progress on the grand challenges facing society. The Inclusive Intelligence Initiative will engage a broad community of scholars from the spectrum of disciplines on campus to put a distinct Berkeley imprimatur on how AI should be conceived, designed, built, and employed. We will extend UC Berkeley's reputation as a leading research center on the technical and societal implications for Artificial Intelligence.

The Inclusive Intelligence (II) Initiative will advance basic research with new theory and algorithms and research and explore ways to shape the design and use of II. The initiative will organize forums for scholarly exchange, artistic creation, and develop new curricula and tools to teach and advise a new generation of interdisciplinary scholars dedicated to applying AI technologies to advance the public good. The initiative will also produce new policy and guidelines so that Berkeley students and alumni will stay at the forefront of this field with a deep understanding of the broader societal issues at stake. The initiative will start by exploring three themes: 1. AI and Society: Supporting Civic Life and Advancing the Public Good. 2. Collaborative AI Systems, Robots, and Livelihoods. 3. AI to Enhance Scientific Discovery.

This report summarizes specific goals, existing units and institutional programs this initiative will build on and new resources needed. This Initiative is synergistic with the new Data Science Division and other Signature Initiatives. This revised report incorporates changes in response to feedback from the Divisional Council (DIVCO), the committees on Academic Planning and Resource Allocation (CAPRA), Diversity, Equity, and Campus Climate (DECC), Graduate Council (GC), and the Undergraduate Council (UGC).

Contents:

Working Group:

Co-Chair: **Ken Goldberg,** Professor, Industrial Engineering and Operations Research
Co-Chair: **Deirdre Mulligan,** Associate Professor, School of Information
Dor Abrahamson, Professor, Graduate School of Education
Joshua Bloom, Professor, Astronomy
Duncan Callaway, Associate Professor, Energy Resources Group, Electrical Engineering and Computer Science

Edmund Campion, Professor, Music
Cathryn Carson, Professor, History
Nicholas de Monchaux, Professor, Architecture, and Director, Berkeley Center for New Media
Marta Gonzalez, Associate Professor, City and Regional Planning
Michael Jordan, Professor, Statistics & Electrical Engineering and Computer Science
Sonia Katyal, Professor, Berkeley Law
Zsolt Katona, Associate Professor, Haas School of Business
Brad de Long, Professor, Economics
Massimo Mazzotti, Professor, History, and Director, Center for Science, Technology, Medicine and Society
Stuart Russell, Professor, Electrical Engineering and Computer Science
Philip Stark, Professor, Statistics

Development Professional: **Jasmine Payne,** College of Engineering
Staff Lead: **Tiff Dressen,** Vice Chancellor for Research Office

Societal challenge:
As Descartes observed, what defines our existence as humans is our ability to think. Today, that ability is undergoing a transformation based on advances in Artificial Intelligence and associated innovations in data science and machine learning that are transforming how we work, learn, communicate, how we govern, and how we live our daily lives.

A fundamental challenge of our time is to ensure that AI positively advances the public good. AI holds enormous potential to empower workers, reduce drudgery, enhance education, accelerate scientific discovery, and connect cultures; it also has shown potential to erode privacy, reduce meaningful work, exacerbate bias and socio-economic disparity, and concentrate power in a few corporations and countries. For example, AI can diagnose skin cancer in images as well as trained doctors and sometimes better. But a widely used AI face recognition system was subject to extreme gender and skin color bias and the intersection of AI and surveillance is adversely affecting human rights in, for example, China. It is vital that AI be designed and integrated in ways that are accurate, democratic, just, safe, and respectful of diverse human values.

At the World Economic Forum in January 2019, Salesforce CEO Mark Benioff announced that AI is advancing far faster than expected, "Those who have AI will be smarter, healthier, richer, and have far more advanced warfare," He characterized AI as a "New Human Right" to be protected. In a similar vein, Russian President Vladimir Putin asserted that whoever leads in artificial intelligence will rule the world. A NYTimes Op-Ed by a leading physician noted that with greater use of AI, there is increased danger that existing income and race-based inequities in the American healthcare system will become invisible and automated and "that we begin to accept the wisdom of machines over the wisdom of our own clinical and moral intuition."

Similar concerns are becoming acute about AI systems perpetuating bias in parole, housing and employment opportunities, and mortgage lending.

UC Berkeley is uniquely positioned as a trusted global leader in a wide range of research on AI, and across the many disciplines where it holds promise to yield new scientific discoveries and innovations in practice. In addition to research in Computer Science, engineering, statistics, scholars from across the humanities, social sciences, and professional schools are using, critiquing, and collaboratively building AI in a manner that is unparalleled at other institutions. As a leading public university with an unequaled legacy pursuing social justice, it is our duty to ensure that AI is not a threat but an opportunity to advance equity, diversity, justice and human flourishing.

How UC Berkeley can address this challenge:
In response, UC Berkeley faculty propose **"Inclusive Intelligence,"** a comprehensive research and teaching agenda for AI that is: 1) Inclusive of diverse disciplines and ways of knowing; 2) Inclusive of diverse populations including those most vulnerable; 3) Inclusive of the spectrum of human and artificial intelligences and the ways they can constructively interact, complement, and enhance each other.

The Inclusive Intelligence Initiative will engage a broad community of scholars from the spectrum of disciplines on campus: engineering, statistics, arts and sciences, law, business, education, liberal arts and humanities, economics, mathematics, physics, material science, astrophysics, chemistry, biology, earth and planetary science, psychology, neuroscience, sociology, political science, history, and philosophy.

There are many open basic research questions about AI and deep learning related to sample complexity, convergence rates, bias-variance tradeoffs, explainability, and covariant shift, the ability of policies to generalize beyond training examples. Basic research on AI theory, algorithms, models, and experiments is very active and many faculty and students in the College of Engineering, Statistics, Mathematics, School of Information, Business, Law, and other departments and units are publishing papers almost daily with basic research results. Many of the world's leading experts in AI and Robotics are affiliated with the Berkeley AI Research (BAIR) group, which includes over 200 faculty and students. Berkeley continues to pioneer basic research in Deep Learning, Unsupervised Learning, Reinforcement Learning, and Algorithmic Fairness. There are equally pressing questions of how best to integrate AI into research to be explored in the biological, physical, and social sciences. The questions of when and how to integrate AI into various domains of public and private life require research from the social sciences, humanities, law, business and other professional domains.

This initiative will advance research with new algorithms, methods, and experiments, and explore how advancements in AI are affecting and affected by individuals, organizations, and publics; organize forums for scholarly exchange; encourage artistic creation that exposes and explores the social and political implications of AI; and teach and advise a new generation of interdisciplinary scholars dedicated to developing, using, governing, and critiquing AI technologies to advance the public good. It will produce new research, curricula, tools, policy, and guidelines so that Berkeley students and alumni can continue to be at the forefront of redefining this field to embrace the broader issues at stake.

This initiative will initially explore three themes:

1. AI and Society: Exploring the Human Experience, Supporting Civic Life and Advancing the Public Good. This theme prioritizes human values and the public good to envision a future where individuals, civic life, and society broadly are supported by intelligent machines and platforms. This theme draws on humanistic traditions to reflect on and interpret the possible transformations AI brings in order to guide them toward the just and humane society in which we hope to live. This theme brings students and researchers together to ask questions such as: How can we contextualize contemporary expectations of AI through literature, philosophy, history, and critical social thought? How can partnerships between AI and humans provide new insights into human values and what is uniquely human? How should we redefine theories of ethics, justice, identity, and responsibility as AI advances? How can AI and robotic systems free humans to focus on and enhance uniquely human skills such as empathy, intuition, and creativity? How could AI be applied to enhance human creativity in the arts and in all academic disciplines? How can expertise in the humanities, social sciences, law, journalism, and the arts be engaged to anticipate and resist potential dangers and drawbacks of AI?

This theme will explore the ramifications of AI for civic life. As cities and public places become "intelligent," what are the opportunities and risks for the citizenry? How might intelligent machines bring us closer together, heighten our sense of collective purpose, and address key risks to vulnerable populations?

Finally, this theme will consider what governance models support inclusively intelligent outcomes—i.e., *yielding the best of human and machine intelligence*. While others have focused on questions of product liability, our agenda will address the myriad of challenges we face as ML/AI is woven into products and services and worksites throughout the economy—from organizational design choices, to standards, to regulatory review and approval to due diligence, testing, and certification. Researchers will explore the design of systems and organizations that preserve, promote, and protect human political and societal values.

The recent online art project (Imagenet Roulette, September 2019) by Berkeley alum Trevor Paglen trained an AI system on the "Person" categories from the ImageNet dataset to expose

racist and sexist labels. This exemplifies how creative work can broaden public awareness about AI.

This theme will bring faculty and students at all levels (undergraduate to graduate and postdoctoral) from across campus together for critical public reflection on the politics of AI systems and their outputs, as well as oversight and accountability. This theme will explore innovations in policy for governing internet-based platforms and content that are consistent with and supportive of democratic norms, human rights and civil liberties.

2. Collaborative AI Systems, Robots, and Livelihoods. Intelligence is not limited to solitary individual minds. Usually it emerges from groups—distributed or local communities, teams, markets, social contracts, and a myriad of methods of communication that expand our collective global intelligence. Increasingly, decisions are made collaboratively as human intelligence is augmented by machines, data, and networks. Incorporating a new "intelligence" with systems that include humans is just as important as trying to emulate, augment, or expand human intelligence. Where AI systems are strong (calculation, memory, vigilance), humans are weak. And vice-versa: where AI systems are weak (empathy, intuition, creativity), humans are strong. New research is needed to develop systems that effectively\ combine human and artificial intelligence. In robotics, there is a need for "co-robotic" systems that combine the perceptual and intuitive skills of humans with the precision of robots (e.g., driver-assist vs. fully autonomous vehicles). Developing collaborative AI systems and robots will require innovations in machine learning, analytical methods, deep learning, computer vision, robotics, cloud computing, and human-machine interaction. This theme will also address the associated policy, economic, and legal issues considering the potential for collaborative AI systems to enhance job satisfaction and human livelihoods by reducing drudgery rather than replacing workers (enhanced safety for Lyft drivers vs. fully autonomous vehicles). Collaborative AI systems have potential to enhance economic efficiency, basic science, worker productivity, morale, and economic stability. This theme will study such systems and design organizational structures, public and organizational policies, and technical standards that support fairness, transparency, contestability, interpretability, and accountability to ensure that the benefits of Collaborative AI are fairly and equitably shared across individuals and groups.

3. AI to Enhance Scientific Discovery. This theme will apply and extend AI to accelerate the scientific process from the design of experiments and the collection of data to scaling inference to accelerate the full spectrum of physical and life sciences. AI may guide us to the next CRISPR, to beyond the Standard Model in Physics, or to an as-yet unimagined solution to climate change. We envision a new landscape of machine learning that is informed and constrained by known physical laws, but also driven by the learning process, leading to discovery of the fundamentals that underlie the natural world. Berkeley can become the world leader for developing novel inference approaches and intuition augmentation in the physical and life sciences, creating lab equipment and control systems that autonomously select in real time

the parameter spaces to explore, and establishing *de facto* research and practices for the modern AI–scientist interface. AI/ML has emerged as a powerful tool for scientific inference; we see it becoming a central substrate of the scientific process itself.

Why Berkeley? UC Berkeley is uniquely positioned as a trusted global leader in AI theory and technology. It is the premier public university, with an unequaled legacy in pursuing social justice and the public good. To tackle this timely set of challenges, there is no institution in the world as qualified, respected, and trusted.

UC Berkeley has been on the forefront of AI since the 1950s and is widely considered the world's leading public university for AI research. Cal professors wrote the most popular textbook for AI, invented several of the most widely used AI and machine learning methods, and have established extensive history of research on AI stretching back to work on cybernetics in the 1950s and 1960s. California Governor Gavin Newsom and state legislators recently invited Cal to advise them on AI and potential impact on workers and the economy.

The breadth of Berkeley faculty engaged with the social, political, economic, and legal implications of AI is unparalleled. Scholars from numerous disciplines are engaged with research both using and studying AI, and often doing so in collaboration with experts in the technology. From the Simon's Institutes programs on privacy and fairness, to the Algorithmic Fairness and Opacity Working Groups ongoing interdisciplinary meetings and speaker series (informed by the earlier Social Science Matrix seminar on "Algorithms as Computation and Culture"), to the Center for Effective Global Action (CEGA), to the CITRIS and the Banatao Institute Berkeley faculty and students are pioneering new ways to design, work with, critique, and govern AI.

As a charter member of the Public Interest Technology University Network — a partnership of 21 colleges and universities dedicated to building the nascent field of public interest technology and growing a new generation of civic-minded technologists — UC Berkeley has staked out its ongoing commitment to educating the next generation of students to more effectively design, build, influence, and govern new technologies in ways that advance the public interest. The II furthers this important joint effort and will solidify the significant role Berkeley plays in producing research and educating a workforce that aligns with the University's public mission and spirit.

Success in 5–10 Years: Inclusive Intelligence will put a distinct Berkeley imprimatur on how AI should be conceived, designed, built, and utilized for the next 100 years. This will extend UC Berkeley's reputation as a leading center for basic research in AI, provide the resources to integrate and bolster the growing community of social science, humanities, and professional school researchers who study AI and its implications, support research necessary for AI to advance scientific inquiry across multiple fields, and guide the field toward inclusivity rather than exclusivity. It is important to act soon: Stanford launched a Human-Centered Artificial

Intelligence (HAI) Institute and MIT launched the Intelligence Quest (IQ) project and a $1B new CS/AI undertaking. Inclusive Intelligence, seeded with private and public funding, can attract major support from major corporations.

Inclusive Intelligence will put UC Berkeley in a global leadership position to advance basic research and define how AI should be conceived, designed, and used for the next century. UC Berkeley's emphasis on Inclusive Intelligence can provide a context and a "conscience" for leaders and practitioners using AI. The Inclusive Intelligence Initiative will develop new technologies—algorithms, and systems, and hardware—research exploring their implications and developing opportunities to shape the design and use of algorithms and methods that align with II, and provide broad-based education to students and guidance to inform public policy. Impact will be measured in terms of students, publications, new courses, patents, startups, and policy proposals for business and government and Inclusive Intelligence systems that are used worldwide.

Examples of Specific Goals for the Three Themes:

1. AI and Society: Exploring the Human Experience, supporting Civic Life and Advancing the Public Good.

A. **Toolkit for for Civic Intelligence.** From autonomous vehicles and other forms of 'smart' civic infrastructure to the effects of social media and the 'gig economy' on urban governance, the systems that govern urban life are increasingly private, closed, proprietary, and opaque to the public. There is a pressing need for public accountability and public values in the governance of cities by and for their citizens. Inclusive AI can build morale and resilience among communities that stand to be most affected. For example, Inclusive AI in transportation can help provide safe, reliable, frequent, and affordable transit service as well as affordable housing and protection for renters. Inclusive AI in energy can engage low-income communities and expose environmental hazards. As the only public university with world-class expertise in fields from public policy to urban planning to the mechanics of autonomous vehicles, UC Berkeley is uniquely positioned to craft an interdisciplinary toolkit for intelligent systems in urban life that can set standards for public accountability and provide a concrete model for the balancing of public and private good in technological and infrastructural systems.
B. **Toolkit for Human and AI Collaboration in the Arts**. The field of AI and the Arts is in its infancy. Several major players, including Google, have launched initiatives in this area, yet the tools and critical approaches related to how AI and artists collaborate remain underexplored. Berkeley is well-situated to explore AI as a creative collaborator with the next generation of artists, choreographers, and composers. Computers can autonomously generate new images and sounds, but humans must judge what is novel or interesting.

C. **Design Principles for Inclusive Intelligence Interfaces.** For humans and machines to work together in alignment with social and political values we must move beyond current research on transparency, explainability, and interpretability. We will study human mental models of intelligent machines and develop a critical assessment of differences between human and machine intelligence, ways of seeing, knowing, and reasoning. We will develop new design principles for human/computer interfaces that convey what AI agents "know" and encourage human engagement and tinkering with machine intelligence (contestability, feedback, experimentation) building on insights from artists and cognitive scientists in the fields of AR/VR.

D. **Novel Governance Models for AI.** As AI is woven into products and services and worksites throughout the economy, these systems will require new legal and policy frameworks for distributing responsibility and liability that encourage mutual engagement between humans and machines. AI embeds values, which must remain visible, consistent with society's expectations, and subject to mechanisms that ensure oversight and accountability. Governance strategies should reinforce the cultural values of autonomy, knowledge, competence, cooperation, and organization, ranging from regulatory review and approval to due diligence to testing and certification. We will build on insights from government regulatory regimes, private-public partnerships, corporate policy, and existing tools while extending and applying these to develop novel governance models for Inclusive Intelligence.

2. Collaborative AI Systems, Robots, and Livelihoods.

A. **Trustable Media Toolkit.** Images, sound and video can be synthesized using "deepfake" software with results that are increasingly difficult to distinguish from reality. We will develop and maintain state-of-the-art tools that allow people to certify, disseminate, and verify images, sound, and video worthy of trust trust. We will develop a collaborative AI system for identifying "deepfakes" and misleading social media posts, analogous to Google's highly successful collaborative tools for filtering spam. This will build on Berkeley expertise as reported in a recent New Yorker article.

B. **Open-Source Markets for Human-to-Human Links**. Many information technology (IT) companies are gathering and processing massive amounts of data for and about humans, ostensibly for the purpose of offering humans better services, but often with the primary purpose of creating an advertising market. It would be transformative if such companies were to focus not on linking the IT company to advertisers but instead on directly linking human producers and consumers. While limited examples of such markets have been created, in domains such as transportation (Uber) and commerce (Ebay and Amazon), they are wholly lacking in domains such as music, journalism, creative writing, and art. Indeed, the IT systems that allow flows of songs, essays, poems and visual art do not attempt to directly link the creators and their audience so that the

creators can profit from their work and so that consumers can shape and influence the creative process. These systems merely attempt to sell more of what is already available. Moreover, they do not take into account real-world scarcity, including limitations on what flesh-and-blood musicians, journalists or artists can produce in real time and at real geographical locations. New ideas will be needed to help industry conceive of and build such systems. For example, recommendation systems, which have been the machine-learning backbone that have brought forth companies such as Amazon, treat each customer independently, and as such they are inadequate in the context of multi-way markets. But it is not clear how to blend machine-learning concepts with the kinds of microeconomic concepts that underlie efficient, healthy multi-way markets, particularly at the scale of modern IT systems. Working with colleagues in industry, we aim to assemble a new discipline that combines core concepts from computer science, economics, and statistics, delivering principles upon which large-scale, human-centered, data-driven markets can be built, with particular focus on domains such as music, journalism, creative writing, and art. We plan to prototype open-source examples of such systems.

C. **Principles for Algorithmic Decision Making in Organizations.** Businesses, governments, and other organizations are making decisions in hiring, admissions, promotions, loans, with the aid of AI algorithms. We will define algorithmic standards that ensure fairness, transparency, interpretability, and accountability; develop tools to improve and automate decision making and to measure the effects; develop metrics to evaluate the quality and fairness of organizational decision making.

D. **Machines as Agents in Markets.** Economies are built on the fundamental idea of efficient markets in which human consumers and producers interact with each other. This initiative explores the role of artificial actors in markets. The emergence of algorithmic agents in financial, media, labor, and other markets often disrupts these in undesired ways, such as disadvantaging human actors, facilitating collusion, or reducing competition. We will investigate how algorithmic agents interact with each other and with humans and explore how they change market outcomes in ways that exacerbate inequality or catastrophic market volatility; design new mechanisms to facilitate interaction between human and algorithmic agents; propose policies and regulations to be inclusive of technological advancements and ensure market efficiency while advancing fairness, equality, and inclusivity for human actors.

E. **Toolkit for Collaborative Causal Policy Analysis.** Inferential statistics, statistical learning, and AI were historically siloed, but we are beginning to understand how to combine them to support causal policy analysis. Consider the problem of absenteeism among K-12 students. Given appropriate data on interventions, statistical learning tools can be used to devise a new policy to counteract absenteeism. Inferential statistics can then be used to predict the impact of the new policy. AI could also help identify situations in which randomized controlled trials are required for rigorous analysis, and

design those trials in ways that are fair to all participants. It is generally believed there are tradeoffs between fairness and accuracy, but empirical experiments have found that both can be simultaneously improved. We will investigate at a sociological and a mathematical level under what conditions fairness considerations can increase causal analysis accuracy. We will develop and openly share a broad suite of tools that support rigorous policy impact analysis on a much larger scale than is possible today.

F. **Inclusive Automation Systems to Enhance Workers.** Rather than replacing workers, AI has potential to enhance workers at all levels. A great deal of technological progress over the past three centuries has taken the form of replacing workers, but an even greater share has taken the form of amplifying the productivity of workers by providing them with better tools. So far the benefits have far outweighed the costs. But, looking forward, we can see a possible future in which large oligopolistic firms focus on developing technology to replace rather than to augment workers. Hence there may well be a very clear need for public universities to do what private oligopolies will not be interested in doing. We will develop Inclusive Automation systems that enhance workers by reducing drudgery in repetitive rote tasks such as inspection on assembly lines, reduce worker fatigue and injuries (eg, exo-skeletons), learn predictive models of human behavior to anticipate and warn of errors; collaborate with workers to learn their capabilities and preferences; enable people to learn, improve, and maintain their expertise, preventing degradation over time. We will start with three specific tasks, warehouse order fulfillment, home and office decluttering, and mechanical assembly. We will incorporating the elements above with novel algorithms that detect errors and facilitate human teleoperation for rapid recovery and continuous learning from negative examples. This research will then be extended to applications in manufacturing, healthcare, logistics, agriculture, and driving.

G. **New Social Contract for the Age of Intelligent Machines**. New public and private principles and policies that take into account potential effects of AI on the livelihoods of workers and income inequality to share the anticipated aggregate economic gains from AI across individuals and groups.

3. AI to Enhance Scientific Discovery

The best chess player in the world today is neither a machine-learning algorithm running atop a computing system nor is it a person: it is a combination of both, working together. Science, we suggest, would be accelerated by approaches and practices that leverage human-software partnerships, imbuing artificial intelligence throughout the scientific process. AI could and should be more than just a tool to analyze data *already* acquired; rather, society would benefit if AI were a first-class actor in data-*taking* decisions, in generating hypotheses, and in developing insights.

A. **Approaches to incorporating physical laws into learning systems**: We envision fundamentally new approaches to ML that can incorporate known "physics" (e.g., conservation laws in physical science). We expect this will lead to faster learning, more accurate results with smaller data sets, better and more relevant predictions, and more rapid understanding of nature and scientific advances. We believe that this research must involve domain experts engaged in answering domain-important questions, working alongside computer scientists and statisticians. Can we employ physics-informed learning to uncover the nature of dark matter and dark energy? The opportunity and challenges are recognized, but there has been essentially no progress. Berkeley has the opportunity and talent to become a world leader.

B. Software and Hardware for Generating and Testing Hypotheses: The Scientific Method involves considering what one or more theories ("models" or "hypotheses") predict in a broad range of circumstances, understanding and imagining definitive experiments that could in principle be performed and the practical issues and costs involved in checking those predictions, then deciding which experiments to perform to winnow the set of competing theories into a smaller set. While this has been until now wholly the domain of *people,* we envision a set of concerted research goals where machines will be involved in the decisions, not just the analyses. Can AI/ML develop and exploit domain-driven intuition? Can AI/ML suggest critical experiments that could lead to scientific revolutions? Can AI/ML find *experimentum crucis*, isolating corners of parameter space where existing theories are likely to break down? We see ripe areas for discovery in life sciences and material science and engineering. For example, can we create models that can suggest new materials to synthesize given a set of desired properties? Success may also help solve the reproducibility crisis in science and the inherent biases in the scientific method.

Examples of Cross-Cutting Goals:

A. **Hiring a cohort of 6-10 new faculty** to enable these cross-cutting research and curricular innovations and to respond to the increase in student demand for these topics.

B. Annual "Inclusive AI" Public Symposium: for scholarly exchange, research, artistic creation, and collective teaching where the changing character of AI-human relations is open to question.

C. New curricula, Majors, Masters programs: Our aim is to train new generations of AI technologists and citizen-practitioners who will have the benefit of Berkeley's intellectual richness, recognizing the multiple dimensions of their own interventions. We want to make Berkeley the place where *all* students have access to these ideas and can incorporate them into their programs of study. We propose to engage faculty across disciplines at the same scale, and with the same integrative intention, as our undergraduate Data Science curriculum -- providing a humanistic foundation for

reflective life and citizenship in a technological world. This curriculum should cross the boundaries from humanistic, artistic, and social scientific to technical, building up to a new undergraduate minor, student research/design opportunities in AI labs, studios, and civil society organizations, graduate courses, and a possible new program in Public Interest Technology. Ultimately, we see the effort percolating through to engage these faculty in the design of the technical curriculum itself. We will develop new curricula so that Berkeley students (and beyond) can be exposed to these ideas and have the opportunity to incorporate them into their programs of study. This curriculum will not only chart new fields of knowledge but also provide essential and original perspectives on the trajectory of AI to date and its possibilities for the future.

D. **AI for Teaching at Scale:** AI-based Software for teaching large scale classes in many disciplines. This will help manage GSIs, graders, assignments, grading, checking for plagiarism, etc.

Participating Units: Inclusive Intelligence is linked closely with the Data Science Commons and the initiatives on Healthcare, Inequality, Democracy, and Public University. Faculty from the following units have expressed interest in leading or participating: Over thirty leading faculty affiliated with the Berkeley AI Research (BAIR) Lab and associated labs in the College of Engineering. The new Division of Data Science and Information has taken it 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, which serves us as a curricular model. Many core machine learning (ML) techniques were developed at UCB, including CART, bagging, boosting, and random forests. Berkeley pioneered major scaling data-analysis projects such as Spark. Berkeley is home to leading experts in AI (Stuart Russell, who wrote the definitive textbook), Computer Vision (Jitendra Malik), Machine Learning (Mike Jordan), Natural Language Processing, and Robotics. Perhaps the most widely used computational tool in Data Science, Jupyter, was developed by UCB faculty member Fernando Pérez. Key methods in numerical analysis, linear algebra, and optimization foundational to machine learning originated at Berkeley (William Kahan, James Demel, Beresford Parlett, Ilan Adler), along with foundational technologies in databases, storage, and even operating systems (Michael Stonebraker, Ken Thompson, Bill Joy, David Patterson, Randy Katz, Garth Gibson). Berkeley's depth in technical foundations is accompanied by an equally strong cohort of faculty from the humanistic and interpretive disciplines who welcome the prospect of engaging with deep challenges around AI, computing, data, and digital connectivity—and with science and engineering colleagues *as* peers. They bring experience and intellectual breadth from high-caliber departments, professional schools, and programs: from Rhetoric, Philosophy, History, and our language departments to Art Practice, Theater, Dance & Performance Studies, Music, and Architecture, to Schools of Law, Information, Policy, Journalism, and Public Health. They have a strong track record of cross-cutting experimentation in New Media, Science & Technology Studies, Digital Humanities, and Design. Berkeley also has a long tradition of

research in economics and social sciences on the impact of technologies on economic performance, politics and policy, and societal health and social cohesion. Inclusive Intelligence will build on ongoing research by UC Berkeley's College of Engineering, Haas Business School, School of Information, School of Law, School of Education, School of Public Policy, and the departments of Statistics, Psychology, Sociology, Political Science, Economics, Philosophy, and History, among others. Faculty leading data-driven and computationally grounded research areas in relation to key societal challenges at the OLab, CEGA, HIFIS, WITS, D-Lab, BCLT, CLTC, FCBA, and CTSP and faculty who created the Algorithmic Fairness and Opacity Working Group (AFOG) have begun tackling key questions set out in the initiative and faculty affiliated with the Berkeley Center for Law and Technology and Samuelson Clinic are tackling key questions of around civil rights, intellectual property, innovation and governance. Faculty from the School of Public Policy, Haas School, Blum Center for Developing Economies, Law, CITRIS, BRIE, UC Berkeley Center for Labor Research and Education, and Granholm project have expressed interest and we expect that many other University projects and units will be engaged, such as neuroscience and the Berkeley Brain Initiative, Work in an Era of Intelligent Tools and Systems (WITS), the Center for Information Technology Research in the Interest of Society (CITRIS), RISELab, the Center for Human Compatible AI (CHAI), the Berkeley Center for New Media, and Berkeley's Arts + Design programs that investigate the impact of artificial and human intelligence in the context of creativity and creative work as well as the Greater Good Science Center, the Townsend Center, LBL, and the Lawrence Hall of Science.

Institutional Programs this Initiative will interact with:

- **Data Science Division:** the "Responsible AI" framework being developed by Jennifer Chayes and Christian Borgs.
- **BAIR:** The Berkeley AI Research Commons
- **Simons Institute**: a premier center for research into theory that is organizing major international workshops on Privacy and Fairness.
- **CITRIS:** The Center for Information Technology Research in the Interest of Society
- **BIDS:** Berkeley Institute for Data Science
- The Center for Science, Technology, Medicine, and Society brings together humanists and social scientists across the Berkeley campus in a concentrated focus on the cultural and political significance of computation, algorithms, and data. It has hosted the Cloud and Crowd working group, administers the Designated Emphasis in Science & Technology Studies, and curates an ongoing Art & Science program with broad public effect
- The Berkeley Center for New Media has provided a home for innovative research, education, and critical engagement across the arts and media, exploring the emerging possibilities of cultural criticism and democratic participation in a digital world.

- The Center for New Music and Audio Technologies in the Department of Music is dedicated to multidisciplinary research with a focus on the creative interaction between music, computers, and emerging technologies.
- The Algorithmic Fairness & Opacity Working Group (AFOG). An interdisciplinary group of UC Berkeley faculty, postdocs, and graduate students developing new ideas, research directions, and policy recommendations around issues of fairness, transparency, interpretability, and accountability in algorithms and algorithm-based systems. These issues require attention from more than just engineers and technologists, as they are playing out in domains of longstanding interest to social scientists and scholars of media, law, and policy, including social equality, civil rights, labor and automation, and the evolution of the news media.
- The Berkeley Center for Law & Technology (BCLT) is a multidisciplinary research center at the University of California, Berkeley, School of Law. The first of its kind, BCLT has garnered worldwide distinction for its research and instructional program exploring the most pressing technology law and policy issues.
- The Center for Effective Global Action (CEGA) is the West Coast hub for global development research. Headquartered at UC Berkeley, CEGA's large, interdisciplinary research network—including a growing number of scholars from low- and middle-income countries—identifies and tests innovations designed to reduce poverty and promote development. Our researchers use rigorous methods as well as novel measurement tools—including wireless sensors, mobile data, and analytics—to evaluate complex programs. Through careful matchmaking, competitive grantmaking, and research dissemination activities, CEGA ensures that the research we produce is relevant, timely, and actionable to policymakers.
- Center for Human Compatible AI (CHAI) is a multi-institution research group based at UC Berkeley, with academic affiliates at a variety of other universities. CHAI's goal is to develop the conceptual and technical wherewithal to reorient the general thrust of AI research towards provably beneficial systems.Artificial intelligence research is concerned with the design of machines capable of intelligent behavior, i.e., behavior likely to be successful in achieving objectives. The long-term outcome of AI research seems likely to include machines that are more capable than humans across a wide range of objectives and environments. This raises a problem of control: given that the solutions developed by such systems are intrinsically unpredictable by humans, it may occur that some such solutions result in negative and perhaps irreversible outcomes for humans. CHAI's goal is to ensure that this eventuality cannot arise, by refocusing AI away from the capability to achieve arbitrary objectives and towards the ability to generate provably beneficial behavior. Because the meaning of beneficial depends on properties of humans, this task inevitably includes elements from the social sciences in addition to AI.

- The D-Lab brings the critical spirit of Berkeley social science and humanities into the training of graduate researchers and postdocs in the methods of data science, machine learning, and AI. Committed to methodological pluralism, to culturally sensitive pedagogy, and to broadening participation for historically excluded identities, D-Lab provides scores of workshops, working groups, and consulting sessions that serve a population of thousands of researchers each year. Students who have passed through its programs have taken positions in academia nationally and internationally, in tech and other industries, and in public interest organizations.
- The Samuelson Law, Technology & Public Policy Clinic is the leading clinical program in technology law and the public interest. Through hands-on, real-world work, the Clinic trains law and graduate students in public interest work on emerging technologies, privacy, intellectual property, free speech, consumer and citizen interests in technology deployment and design, creativity, innovation, and other information policy issues. The Clinic was established in January 2001 and was the first in the nation to provide students with the opportunity to represent the public interest in sound technology policy through client advocacy and participation in legislative, regulatory, litigation and technical standard setting activities. Today, the Samuelson Clinic functions as both a traditional legal Clinic and as a site of interdisciplinary, policy-relevant research.
- The Human Rights Investigations Lab—the first university-based initiative of its kind in the world—engages more than 80 graduate and undergraduate students from multiple disciplines in real-world open source investigations of human rights abuses and potential war crimes. Part of UC Berkeley's Human Rights Center, which works at the intersection of technology, law, and human rights, the lab trains students on cutting-edge tools and methods for analyzing and verifying social media content in order to make the information and potential evidence more usable by courts, international NGOs, and media. We're creating a pipeline of trained students to enter this new field. The Human Rights Center is also working with global leaders in technology and law to draft the first international protocol on open source evidence.
- The Citizen Clinic the world's first public-interest cybersecurity clinic supports the capacity of politically-vulnerable organizations and communities —including media outlets, human rights groups, and non-government organizations that are at risk of cyberattack -- to defend themselves against online threats. Developed by the Center for Long-Term Cybersecurity at UC Berkeley, the Clinic builds off CLTC research on the ecosystem providing technical assistance to civil society, the Clinic supports interdisciplinary teams of students to assess threats to targeted communities, recommend risk-appropriate mitigations, and to work collaboratively with partner organizations to implement new policies and technical controls that enhance their cybersecurity. The Clinic is now in its third iteration, having seen growth in both its number of students and supported clients by 50% each semester.

- The Center for Technology, Society & Policy (CTSP) is a multidisciplinary research & design/build center, based in the School of Information and co-directed by two PhD-students, focusing on the emergent social and policy issues arising from the development and adoption of technology. Since 2015, CTSP has provided nearly 100 small fellowships to graduate students and postdocs to develop an on-ramp to public interest technology work and has developed programming to cultivate the public interest technologist community (from encryption events to help students learn about interpersonal cybersecurity best practices to an event for non-profits and city partners to pitch students on projects and a data for good competition with judging from industry and government). CTSP supported projects are at the intersection of social interest and policy concerns, whether centering privacy, fairness, or inclusion, or looking at new uses of algorithms, blockchain technologies, genetic technology, or drones.
- Graduates for Engaged and Extended Scholarship in Computing and Engineering (GEESE) Geese's mission is to develop community among graduate students and postdocs interested in working at the intersection between engineering and the social sciences and humanities. Inspired by Berkeley's long history of advocating for social progress, we hope to create a vibrant transdisciplinary network of future scholars prepared to tackle the emerging ethical dilemmas and opportunities for social change presented by new forms of technological disruption.
- Undergraduates for Human Contexts and Ethics of Data (working on their website) are engaging students across the disciplines to explore the societal and human questions raised by data science, machine learning, and AI. With the support of the Division of Data Sciences, they are organizing reading groups and symposia, creating a blog and Medium outlet, and mobilizing to support STEM faculty in integrating HCE content into their courses.

Resources To Build On

This initiative will build on existing resources but require significant additional time investments by faculty and students. In the short run, funding for a major public symposium and seed grants can motivate faculty to engage in new research projects and plan new courses.

1. New FTE: We estimate: 6-10 additional faculty positions in strategic areas. Such growth is aligned with campus plans to add 100 new faculty to the University and would allow us to meet student demand across departments and programs. We have the opportunity to create faculty clusters across departments and to create joint positions in diverse subjects, such as Gender Studies and Computer Science, or Law and Data Science. Such new positions would support the cross-cutting and interconnected research, teaching, and societal impact we seek to enable through this initiative.
2. Seed Grants: Faculty Innovation Awards can provide seed grants to Berkeley faculty through a competitive application process. Faculty members and teams will submit

proposals for funding, and a panel of Berkeley faculty members and key campus leaders will review proposals and allocate the awards.

3. Seed Grants and teaching release to incentivize new curricula: A key goal of the Inclusive Intelligence strategic initiative is to deepen the engagement of Berkeley students with the broad questions and potential implications of the rise of AI in our daily lives—encompassing a variety of fields. Through a grant program open to Engineering faculty, the Initiative would support faculty as they design novel courses to address the topics in the initiative, with priority to courses that are team-taught by professors in different disciplines, or which combine technical content with humanistic or social science subjects. Such a program would also enable the campus to update existing courses—utilizing new technologies in the classroom, or incorporating themes and content in within the initiative in new ways. Course Innovation funds could provide grants to faculty members who will create new curricula and expand the reach of existing courses.
4. Graduate Student and Visiting Scholar Fellowships: many PhD students are actively engaged in research on AI and every aspect of the research proposed for this initiative would involve graduate students (as well as undergraduate, masters, and postdoctoral students). Graduate students from diverse disciplines are engaged in research on the cross-cutting issues of fairness, transparency, interpretability, and accountability in algorithms and algorithm-based systems. For this reason, communities supported by AFOG and CHAI provide important intellectual homes to a diverse community of PhD students and postdoctoral scholars from across campus. A very important goal of the Inclusive Intelligence Initiative is to provide more resources in terms of space, facilities, courses, and funding for graduate students, and interdisciplinary research activities necessary to support the various forms of inclusivity.
5. Support for Undergraduates: There is a huge and growing demand from undergrads for more courses and research opportunities in AI writ large. The interest is in core technical fields, as well as the humanities and social sciences. We would like to develop several Big Ideas undergraduate classes. As Data Science is closely related to AI another goal is to greatly expand the number and capacity of undergraduate courses on AI and robotics, and expand the number of summer and part-time research opportunities for undergraduates
6. Physical space in which academic communities related to each theme can emerge.
7. Support for academic and public forums, exhibitions, and conferences; faculty and student programs, such as residencies and fellowships for extended study. A strong, stable institutional home is needed in the space in between academic departments, integrally networked across engineering and the arts, humanities, and social sciences and porous to flows from all corners of campus.