RESPONSIBLE ARTIFICIAL INTELLIGENCE //

Recommendations to Guide the University of California's Artificial Intelligence Strategy

University of California Presidential Working Group on AI
FINAL REPORT | OCTOBER 2021

UNIVERSITY OF CALIFORNIA

TABLE OF CONTENTS

ABOUT

The University of California (UC) is increasingly incorporating artificial intelligence as a means to improve its operations. While AI can bring significant benefits, ill-conceived deployments risk imposing disproportionate harms. In order to guide the appropriate oversight and implementation of AI, UC President Michael Drake and former UC President Janet Napolitano formed a Presidential Working Group charged with developing overarching principles and recommendations for UC's current and future use of AI.

Launched in August 2020, the interdisciplinary "UC Presidential Working Group on Artificial Intelligence" (hereinafter "Working Group") is composed of 32 faculty and staff from all 10 UC campuses, as well as representatives from UC Legal; Office of Ethics, Compliance and Audit Services (ECAS); Procurement; Office of the Chief Information Officer; Research & Innovation; and UC Health, among others. The members represent diverse disciplines, including computer science and engineering, law and policy, medicine, and the social sciences. The Working Group met monthly from August 2020 to August 2021 and developed a set of UC Responsible AI Principles to guide the procurement, development, implementation, and monitoring of AI as may be implemented by UC. The Working Group does not address issues related to academic research on AI. Instead, the Working Group is focused on providing guidance to UC on appropriate implementation of AI in service provision (e.g., in the administration of human resources services and UC Health services).

To provide guidance on strategies to operationalize the UC Responsible AI Principles, the Working Group established four subcommittees, each corresponding to a high-risk AI application area: health, human resources (HR), policing, and student experience. Each subcommittee was tasked with conducting research on how AI is or will be used in its application area, summarizing the benefits and risks of application, and providing guidance on how UC should operationalize the UC Responsible AI Principles to mitigate harms and maximize benefits of AI.

The Working Group sought to provide President Michael Drake with guidance to:

- Develop methods and mechanisms to operationalize the UC Responsible AI Principles in the use of existing AI and the development of new applications of AI within UC service provision, especially in areas likely to affect individual rights, including health, HR, policing, and student experience.

- Make further recommendations on appropriate data stewardship standards for UC data that may be used in the development and use of AI-enabled tools and systems.

- Create the foundation for coordinated campus-level councils and systemwide coordination through the University of California Office of the President (UCOP) that will further the principles and guidance developed by this Working Group to counter the potentially harmful effects of AI and strengthen positive outcomes within UC services.

The report concludes with four primary recommendations that seek to achieve these goals.

MEMBERS OF THE UC AI WORKING GROUP

Working Group Co-Chairs
Alexander Bustamante, JD, SVP Chief Compliance and Audit Officer, UCOP
Brandie M. Nonnecke, PhD, Director, CITRIS Policy Lab, UC Berkeley
Stuart Russell, PhD, Professor, UC Berkeley

Health Subcommittee
Co-Chair Cora Han, Chief Health Data Officer, UC Health, UCOP
Co-Chair Barbara A. Koenig, Director, UCSF Program in Bioethics
Co-Chair Jessica Newman, Research Fellow, Center for Long-Term Cybersecurity, UC Berkeley
Ken Goldberg, Professor, Industrial Engineering and Operations Research, UC Berkeley
Hillary Noll Kalay, Senior Counsel, Health Affairs and Technology Law, UC Legal
Sonia Katyal, Professor, Berkeley Law
Yang Liu, Assistant Professor, Computer Science and Engineering, UC Santa Cruz
Camille Nebeker, Associate Professor, UC San Diego Design Lab and Wertheim School of Public Health, UC San Diego
Lawrence Saul, Professor, Department of Computer Science and Engineering, UC San Diego

Human Resources Subcommittee
Co-Chair Mark Cianca, Chief Information Officer, UCOP
Co-Chair Alexa Koenig, Executive Director, Human Rights Center, UC Berkeley
Yufei Ding, Assistant Professor, Dept. of Electrical and Computer Engineering, UC Santa Barbara
Mary Gauvain, Distinguished Professor, Psychology, UC Riverside and Chair, UC Academic Senate
Bill Maurer, Dean, School of Social Sciences and Professor, Dept. of Anthropology, UC Irvine
Mark Nitzberg, Executive Director, Center for Human-Compatible AI, UC Berkeley
Hoyt Sze, Managing Counsel, Health Law Group, UC Legal

Policing Subcommittee
Co-Chair Hany Farid, Professor, Electrical Engineering and Computer Science and School of Information, UC Berkeley
Co-Chair Safiya Noble, Associate Professor, Dept. of Information Studies, UCLA
Kevin Confetti, Deputy Chief Risk Officer, UCOP
Nadia Heninger, Associate Professor, Computer Science and Engineering, UC San Diego
Elizabeth Joh, Professor, UC Davis School of Law
Deirdre K. Mulligan, Professor, School of Information, UC Berkeley

Student Experience Subcommittee
Co-Chair Camille Crittenden, Executive Director, CITRIS and the Banatao Institute
Co-Chair Liv Hassett, Associate Campus Counsel, UC Berkeley
Miguel Á. Carreira-Perpiñán, Professor, Electrical Engineering and Computer Science, UC Merced
Andrew W. Houston, Counsel, Business, Finance, and Innovation Practice Group, UC Legal
Shanda Hunt, Systemwide Research Compliance Officer, Ethics, Compliance and Audit Services, UCOP
Caroline Jeanmaire, Director of Strategic Research and Partnerships, CHAI, UC Berkeley
Theresa Maldonado, Vice President for Research and Innovation, UCOP

Special thanks to CITRIS Policy Lab Research Assistants Shalin Brahmbhatt and Gurbir Singh and Alyssa Ivancevich at UC Hastings for their careful review and to Brandie Nonnecke for leading the writing and design of this report.

GLOSSARY

Artificial Intelligence
Technologies that aim to reproduce or exceed abilities in computational systems that would require human-like thinking to perform a wide range of tasks, from simple to sophisticated.[1]

> An ***AI system*** is a machine-based system that is capable of influencing the environment by making recommendations, predictions, or decisions for a given set of objectives. It uses machine and/or human-based inputs/data to: i) perceive environments; ii) abstract these perceptions into models; and iii) interpret the models to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy.[2]
>
> ***AI-enabled tools*** refer to AI systems that are implemented within a particular context (e.g., a natural language processing chatbot used by a UC campus to respond to students' inquiries about the admissions process).

Automated Decision System (ADS)
ADS can be defined as "a computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision or facilitates human decision making."[3]

Automatic License Plate Readers (ALPR)
ALPRs are high-speed, computer-controlled camera systems that automatically capture all license plate numbers that come into view, along with data on location, date, and time.[4]

Computer Vision
A field of artificial intelligence that trains computers to interpret and understand the visual world.[5]

Deep Learning
A form of machine learning that uses neural networks to process data. It is well suited to tackle tasks involving complex data such as images, video, sound files, and unstructured text. However, it can be difficult to explain how deep learning algorithms arrived at a decision.[6]

Explainability
Explainability of a machine learning model refers to how easy it is to understand the internal logic the model uses to make a prediction. Linear models (such as logistic regression) and small decision

---

[1] "AI Procurement in a Box: AI Government Procurement Guidelines," World Economic Forum, June 2020, http://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf.

[2] Karine Perset et al., "A first look at the OECD's Framework for the Classification of AI Systems, designed to give policymakers clarity," *The AI Wonk* (blog), *OECD.AI Policy Observatory*, November 24, 2020, https://oecd.ai/wonk/a-first-look-at-the-oecds-framework-for-the-classification-of-ai-systems-for-policymakers.

[3] Personal Rights: Automated Decision Systems, A.B. 2269, 2019-2020 Reg. Sess. (Cal. 2020), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB2269.

[4] "Automated License Plate Readers (ALPRs)," *Electronic Frontier Foundation*, Aug. 28, 2017, https://www.eff.org/pages/automated-license-plate-readers-alpr.

[5] "An Executive's Guide To Real-World AI: Lessons from the Front Lines of Business," Harvard Business Review Analytic Services, 2019, https://enterprisersproject.com/sites/default/files/an_executives_guide_to_real_world_ai.pdf

[6] Ben Dickson, "What is Deep Learning?" *PCMag*, last modified August 8, 2019, https://www.pcmag.com/news/what-is-deep-learning.

trees are on the more explainable end of the spectrum; neural nets and decision forests are on the less explainable end (often referred to as "black-box").

Facial Recognition
A method of biometric identification to verify the identity of a person through facial patterns.[7]

Machine Learning (ML)
ML is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.[8]

Natural Language Processing (NLP)
A computer's attempt to understand, interpret, and manipulate spoken or written language, which often involves machine learning. It must parse vocabulary, grammar, and intent, and allow for variation in language use.[9]

Neural Networks
Interconnected layers of software-based calculators known as "neurons" form a neural network. In a neural network, a set of units receives pieces of data, for example pixels in a photo, performs simple computations on them, and passes the results on to the next layer of units, eventually reaching the answer in the final layer. Neural networks are a part of deep learning.[10]

Robotic Process Automation (RPA) Bots
RPA is software "robots" that can automate rules-based tasks. The bots automate tasks such as processing transactions, manipulating data, responding to queries, and communicating with other systems. Not all RPA bots use AI. AI can be implemented into RPA to process and gather insights from semi- and unstructured data to structure RPA processes.[11]

Supervised Learning
Supervised learning occurs when the given output and input variables are provided to the algorithm and the algorithm uses training data combined with human guidance to learn the relationship between the given inputs and the given output.[12]

Unsupervised Learning

---

[7] "Facial Recognition: How it Works and its Safety," *Electronic Identification*, July 2021, https://www.electronicid.eu/en/blog/post/face-recognition/en.

[8] "Machine Learning: What it is and Why It Matters," *SAS*, accessed October 30, 2020, https://www.sas.com/en_us/insights/analytics/machine-learning.html.

[9] Matthew Hutson, "AI Glossary: Artificial Intelligence, in So Many Words," *Science* 357, no. 6346 (July 2017):19, http://doi.org/10.1126/science.357.6346.19.

[10] "An Executive's Guide to AI," *McKinsey & Company*, accessed October 30, 2020, https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/An%20executives%20guide%20to%20AI/An-executives-guide-to-AI.ashx;
Matthew Hutson, "AI Glossary: Artificial Intelligence, in So Many Words," *Science* 357, no. 6346 (July 2017):19, http://doi.org/10.1126/science.357.6346.19.

[11] "An Executive's Guide To Real-World AI: Lessons from the Front Lines of Business," *Harvard Business Review Analytic Services*, 2019, https://enterprisersproject.com/sites/default/files/an_executives_guide_to_real_world_ai.pdf.

[12] "An Executive's Guide to AI," *McKinsey & Company*, accessed October 30, 2020, https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/An%20executives%20guide%20to%20AI/An-executives-guide-to-AI.ashx.

Unsupervised learning is when an algorithm explores and identifies patterns in the given input data without being provided an explicit input or output variable.[13]

EXECUTIVE SUMMARY

The University of California (UC) is increasingly turning to AI-enabled tools to support greater efficiency, effectiveness, and equity of its operations. Use of AI can assist in the provision of healthcare within UC Health, in the identification of qualified candidates for jobs and promotions, in the appropriate use of policing on campus, and in the administration of innovative teaching and assessment methods. While AI can bring significant gains, ill-conceived deployments hold great risk, such as producing biased or discriminatory practices.

The UC Presidential Working Group on Artificial Intelligence was launched in 2020 by UC President Michael Drake and former UC President Janet Napolitano to assist UC in determining a set of responsible principles to help guide its procurement, development, implementation, and monitoring of AI in its delivery of services. The Working Group does not address issues related to academic research on AI. Instead, it is focused on providing guidance to UC on appropriate oversight and implementation of AI in service provision[14]

While AI holds great potential, UC must carefully consider whether and how to implement it, how to identify and mitigate potential harms, and discover appropriate strategies to maximize benefits. To support this work, the Working Group conducted interviews with dozens of experts and stakeholders across UC and administered a survey to campus chief information officers (CIOs) and chief technology officers (CTOs) to better understand how AI is implemented and the governance and oversight mechanisms in place. Results from the interviews and survey indicate an overwhelming demand for oversight mechanisms during the procurement process and the need for UC-wide strategies and guidance.

To support these goals, the Working Group developed a set of UC Responsible AI Principles and explored four high-risk application areas: health, human resources, policing, and student experience. This report explores current and future applications of AI in these areas and recommendations for how to operationalize the UC Responsible AI Principles. The report concludes with overarching recommendations to help guide UC's strategy for determining whether and how to appropriately implement AI in its service provision.

Drawing upon insights gleaned from the Working Group's literature review, interviews with domain experts and stakeholders, the CIO/CTO survey, and the work of the four subcommittees, we propose the following overarching recommendations:

1. **Institutionalize the UC Responsible AI Principles** in procurement and oversight practices;
2. **Establish campus-level councils and systemwide coordination through the University of California Office of the President (UCOP)** that will further the principles and guidance developed by this Working Group;

[13] Ibid.

[14] For example, in the administration of human resources services and UC Health services. Research may fall under the guidance and recommendations of this report if the research is used to inform the development and/or use of AI-enabled tools employed in the provision of services within UC.

3. **Develop a risk and impact assessment strategy** to evaluate AI-enabled technologies during procurement and throughout the tool's operational lifetime, including recommendations for appropriate risk-mitigation strategies; and
4. **Document AI-enabled technologies in a public database** to promote transparency and accountability.

INTRODUCTION

Artificial intelligence (AI) has the potential to transform UC's operations.[15] Its application can improve the quality of decision-making, enable greater efficiency in core university functions, and better ensure equity in service delivery.[16] While AI holds potential, it simultaneously poses ethical, privacy, safety, equity, and security risks. Inappropriate, inaccurate, or inconsistent data and ill-considered assumptions in model design can lead to problematic outcomes, such as biased or discriminatory decisions. UC has put in place robust policies and guidelines for technology procurement and use, especially related to data privacy and security.[17] However, AI-driven processes present additional challenges due to their scale, velocity, opacity, and potential for limited human oversight.

To better understand how AI is currently used within UC and oversight mechanisms in place, the Working Group interviewed experts and community stakeholders across UC, including faculty, students, and staff representatives from UC Legal; Office of Ethics, Compliance and Audit Services (ECAS); Procurement; Office of the Chief Information Officer, among others. A survey was administered to campus chief information officers and chief technology officers.[18] Results indicate an overwhelming concern for the potential risks of AI-enabled tools, especially risks of harm related to bias and discrimination and the need to implement appropriate oversight mechanisms, especially during procurement because most AI-enabled tools used within UC operations are procured from third-party vendors. In open-ended comments, respondents urged UC to provide systemwide guidance on strategies to identify and evaluate risks of AI-enabled tools and appropriate governance mechanisms, while allowing for flexibility at the campus level to conform to local policies and norms.

In light of the concerns raised during interviews and the survey, the Working Group developed a set of UC Responsible AI Principles to help guide UC's appropriate procurement, development, implementation, and monitoring of AI. To elucidate these principles, the Working Group established four subcommittees to explore current and potential use of AI within four areas that pose high risk to individual rights, including: health, human resources (HR), policing, and student experience. Each subcommittee explored current and potential uses of AI within their application area and operationalized the UC Responsible AI Principles that are particularly salient to each use case. The

[15] Artificial intelligence encompasses a wide range of technologies embodied in algorithms that process information to make decisions or control systems. This report primarily focuses on the use of AI-enabled tools, such as machine learning (ML) algorithms and automation tools such as robotic process automation (RPA) and chatbots.

[16] While AI can support equity in service delivery, for example by revealing discriminatory or biased decision-making, it is value-laden. The values and prioritizations of the programmers and implementers shape how concepts such as fairness and non-discrimination are defined and operationalized. This process itself is not immune to bias.

[17] "UC Procurement: UC Systemwide Legal Documents and Policies," *University of California Office of the President*, https://www.ucop.edu/procurement-services/policies-forms/index.html

[18] The survey was completed by CIOs and CTOs at the following campuses: UC Berkeley, UCLA, UCLA Health, UC Merced, UC Riverside, UC Santa Barbara, UC Santa Cruz, UC San Diego, UC San Diego Health, and University California Office of the President (UCOP). The implementation of the survey was not considered formal research and was not reviewed by an Institutional Review Board.

report concludes with overarching recommendations to guide UC's responsible implementation of AI-enabled tools. An appendix containing a non-exhaustive list of relevant laws, regulations, and university policies that are likely to affect UC's development and use of AI within health, HR, policing, and student experience is provided. Legal counsel on each campus should be consulted.

UC RESPONSIBLE AI PRINCIPLES

In response to growing concerns over the use and consequences of AI, at least 170 sets of AI principles and guidelines have been developed to guide the private and public sectors' responsible development and use of AI.[19] While the sets of principles vary in style and scope, a consensus is growing around key themes, including the need for accountability, privacy and security, transparency and explainability, fairness and non-discrimination, professional responsibility, human control, and the promotion of human values like civil and human rights.[20]

Drawing upon these common themes and from insights gleaned from the deep expertise of the members of the UC Presidential Working Group on Artificial Intelligence, the Working Group urges UC to adopt the following responsible AI principles to guide its procurement, development, implementation, and monitoring of AI within its provision of services:

1. **Appropriateness:** The potential benefits and risks of AI and the needs and priorities of those affected should be carefully evaluated to determine whether AI should be applied or prohibited.

2. **Transparency**: Individuals should be informed when AI-enabled tools are being used. The methods should be explainable, to the extent possible, and individuals should be able to understand AI-based outcomes, ways to challenge them, and meaningful remedies to address any harms caused.

3. **Accuracy, Reliability, and Safety:** AI-enabled tools should be effective, accurate, and reliable for the intended use and verifiably safe and secure throughout their lifetime.

4. **Fairness and Non-Discrimination:** AI-enabled tools should be assessed for bias and discrimination. Procedures should be put in place to proactively identify, mitigate, and remedy these harms.

5. **Privacy and Security**: AI-enabled tools should be designed in ways that maximize privacy and security of persons and personal data.

6. **Human Values:** AI-enabled tools should be developed and used in ways that support the ideals of human values, such as human agency and dignity, and respect for civil and human rights. Adherence to civil rights laws and human rights principles must be examined in consideration of AI-adoption where rights could be violated.

7. **Shared Benefit and Prosperity:** AI-enabled tools should be inclusive and promote equitable benefits (e.g., social, economic, environmental) for all.

8. **Accountability:** The University of California should be held accountable for its development and use of AI systems in service provision in line with the above principles.

---

[19] "AI Ethics Guidelines Global Inventory," *Algorithm Watch*, accessed October 27, 2020, https://inventory.algorithmwatch.org/.

[20] Jessica Fjeld et al., "Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI," *Berkman Klein Center for Internet & Society, Harvard University*, 2020, https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf.

AI USE CASES & SUB-RECOMMENDATIONS

The Working Group examined four application areas that pose high risk to individual rights including health, human resources, policing, and student experience. This section provides an overview of current and potential uses of AI in each area, explores the benefits and risks of implementation, and provides recommendations for how to operationalize the UC Responsible AI Principles.

Introduction

The UC AI Working Group formed a health subcommittee due to the increasing interest in and impact from AI-enabled tools in connection with UC health data and services. The subcommittee examined the benefits and risks arising from current and potential uses of AI in healthcare, and devised several recommendations intended to mitigate these risks, promote use ofAI in an equitable and responsible manner, and support inquiry to explore the broader social implications of AI within healthcare. The recommendations are intended to help UC operationalize the UC Responsible AI Principles throughout its health services. In addition to discussions among subcommittee members, the group consulted with relevant external practitioners and academics and sought input from many UC experts on AI and health.[21]

The use of AI in UC Health requires attention because advances in AI have the potential to revolutionize the practice of medicine and healthcare. What might we expect when AI-enabled tools interact with patient records, family histories, X-rays and CT scans, genetic profiles, wearable devices, and the full text of thousands of medical journals? With ever-increasing stores of data, AI-based approaches have the potential to usher in a new era of personalized healthcare and precision medicine. In many specialties, these approaches have already yielded better automated tools for the detection and/or prevention of disease.[22] AI-enabled tools are also increasingly improving

[21] Experts within UC included: Atul Butte, Chief Data Scientist at UC Health and Director of the Bakar Computational Health Sciences Institute at UCSF; Ziad Obermeyer, Associate Professor and Blue Cross of California Distinguished Professor of Health Policy and Management at the Berkeley School of Public Health; Laurel Riek, Professor of Computer Science and Engineering at UC San Diego; Rachael Callcut, Vice Chair Clinical Science and Division Chief, Trauma, Acute Care Surgery, and Surgical Critical Care, Naveen Raja, Medical Director of Population Health and Associate Clinical Professor, David Geffen School of Medicine at UCLA; UC CIOs and CTOs, and members of health data governance groups from UCI Health, UCSD Health, UCLA Health, UCSF Health, and UCD Health.

[22] Rushabh Shah and Alina Chircu, "IoT and AI in healthcare: A systematic literature review," *Issues in Information Systems*, Volume 19, Issue 3, pp. 33-41, 2018, https://iacis.org/iis/2018/3_iis_2018_33-41.pdf.

systemic aspects of healthcare in ways as simple as scheduling follow-up appointments or as consequential as stemming large-scale fraud.

The number of scientific publications related to AI in health has grown steadily and significantly in recent years,[23] and the COVID-19 pandemic has increased the prevalence of AI-powered health solutions around the world.[24] At best, AI-enabled tools provide opportunities to make UC's practice and delivery of healthcare more effective, efficient, and equitable. For example, as described in more detail below, AI-enabled tools can improve diagnostic technologies, identify workflow optimization strategies, and aid in quality and population health efforts.[25] AI-enabled tools have also improved healthcare delivery throughout the COVID-19 pandemic in varied ways, including helping track and predict outbreaks, countering misinformation, and supporting biomedical and pharmacotherapy studies.[26]

Notwithstanding the tremendous potential for AI in this domain, it is necessary to proceed with caution. There are enormous inequities in our current health-care system, inequities that may be exacerbated if AI is based on datasets that are biased or incomplete.[27] In addition, many current approaches to AI work in ways that are inscrutable to physicians and patients alike. Such approaches lack the transparency and accountability that must be part of any ethical framework for high-stakes decision-making, and particularly for a public research university.[28] Finally, the benefits of AI (and machine learning, in particular) must be weighed against the loss of privacy from large-scale data collection. The sensitive nature of health data as well as the impact of health-related decisions on people's lives and wellbeing mean that mitigating risks, and in some cases avoiding certain AI uses that may present significant ethical, legal, privacy, and/or security risks that cannot be mitigated must be a priority. UC is at the forefront of efforts to design and implement AI technologies into medicine and healthcare and must take these considerations into account.

Many of the risks associated with AI, including privacy and security vulnerabilities, racial and gender bias, and the lack of explainability of deep learning models built using current methods, are generic to all AI uses. Other considerations primarily related to the collection, privacy, and security of health data, including the risks of re-identification of data, are featured in other reports.[29] However, we highlight here several examples specific to AI in the healthcare domain. For example, a widely used algorithm used to predict patients' health risk was found to be racially biased and underserving Black patients because it used healthcare cost as a proxy for illness without

---

[23] Hashiguchi, Tiago Cravo Oliveira, Luke Slawomirski, and Jillian Oderkirk. "Laying the foundations for artificial intelligence in health." (2021), https://www.oecd-ilibrary.org/docserver/3f62817d-en.pdf.

[24] AAAS. A*rtificial intelligence and COVID-19: Applications and impact assessment*, May 2021. Report prepared by Ilana Harrus and Jessica Wyndham under the auspices of the AAAS Scientific Responsibility, Human Rights and Law Program, https://www.aaas.org/sites/default/files/2021-05/AIandCOVID19_2021_FINAL.pdf.

[25] Michael Matheny et al., "Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril." *NAM Special Publication. Washington, DC: National Academy of Medicine* (2019): 154, https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf.

[26] AAAS. A*rtificial intelligence and COVID-19: Applications and impact assessment*, May 2021. Report prepared by Ilana Harrus and Jessica Wyndham under the auspices of the AAAS Scientific Responsibility, Human Rights and Law Program, https://www.aaas.org/sites/default/files/2021-05/AIandCOVID19_2021_FINAL.pdf.

[27] Camille Nebeker, John Torous, and Rebecca J. Bartlett Ellis. "Building the case for actionable ethics in digital health research supported by artificial intelligence." *BMC medicine* 17, no. 1 (2019): 1-7.

[28] Ibid.

[29] "Report to the President: President's Ad Hoc Task Force on Health Data Governance," *University of California,* Jan. 26, 2018, https://www.ucop.edu/uc-health/reports-resources/health-data-governance-task-force-report.pdf.

understanding the context of divergent healthcare spending across communities.[30] Another important example comes from a study of AI-based COVID-19 diagnosis studies, which revealed that out of 62 studies reviewed for potential clinical utility, none were found to be usable due to methodological flaws and/or underlying biases.[31] An additional important area of consideration for healthcare is the persistent challenges and biases associated with AI language models that make use of natural language processing (NLP), especially given that NLP is being used for medical chatbots working directly with patients.[32]

The use of AI for health not only poses risks in clinical settings, but also at a structural level due to shifts in power, control, and access to care. For example, the World Health Organization (WHO) warns that AI could, "subordinate the rights and interests of patients and communities to the powerful commercial interests of technology companies or the interests of governments in surveillance and social control."[33] Mitigating these risks will require appropriate governance and oversight over the use of AI in healthcare.

"The use of AI for health not only poses risks in clinical settings, but also at a structural level due to shifts in power, control, and access to care."

It is important to note that for now, the use of AI in healthcare—especially in clinical settings—is still limited.[34] In many cases, low-tech or no-tech solutions are a better fit for persistent problems. Where AI technologies are used in clinical settings, they typically provide support to healthcare professionals on extremely narrow tasks. Full automation of healthcare services remains distant and may never be appropriate given persistent technical and social barriers. For example, commonly used healthcare algorithms are still often incorrect and require regular oversight.[35] A WHO report on AI in health released June 2021 states that while AI could help empower patients to take better control of their healthcare and provide rural communities with greater access to care, people should not overestimate the benefits of AI for health given persistent challenges related to collecting health data, mitigating bias, and risks of AI to patient safety, cybersecurity, and the environment.[36]

Despite the challenges enumerated above, AI remains a powerful tool that—with appropriate governance and oversight—may be able to help the University of California Health (UCH) deliver critical services. Several key categories of use cases most relevant to UCH are described below, including clinical diagnosis and treatment, workflow improvement, and population health

---

[30] Ziad Obermeyer et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science. 366.* 447-453. 10.1126/science.aax2342.

[31] Michael Roberts et al. "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans." *Nature Machine Intelligence* 3, no. 3 (2021): 199-217.

[32] Emily Bender et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *ACM FAccT* (2021): 610-623. https://dl.acm.org/doi/pdf/10.1145/3442188.3445922.

[33] "Ethics and Governance of Artificial Intelligence for Health." *World Health Organization*, June 28, 2021, https://www.who.int/publications/i/item/9789240029200.

[34] Michael Matheny et al., "Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril." *NAM Special Publication. Washington, DC: National Academy of Medicine* (2019): 154, https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf.

[35] Tom Simonite. "An Algorithm That Predicts Deadly Infections is Often Flawed." *Wired*, June 21, 2021, https://www.wired.com/story/algorithm-predicts-deadly-infections-often-flawed.

[36] "Ethics and Governance of Artificial Intelligence for Health." *World Health Organization*, June 28, 2021, https://www.who.int/publications/i/item/9789240029200.

improvement. The successful use of AI in the health domain will be aided by the consistent implementation of the UC Responsible AI Principles and the recommendations set forth in this report.

The actions that UCH takes concerning the use of AI matter because UCH plays a significant role in the US medical system. UCH comprises six academic health centers, 20 health professional schools, a Global Health Institute, and systemwide services that improve the health of students, faculty, and employees.[37] In Fiscal Year 2019-20, UCH provided approximately 8.1 million outpatient visits and 1.1 million inpatient days. In addition, UCH is one of the largest providers of care to Medi-Cal enrollees despite representing less than 6% of the 74,180 non-federal, short-term, acute care hospital beds in California.[38]

Use Cases for AI in Health

Many current and future contemplated uses of AI in healthcare across UC could affect clinicians, administrators, patients, and their families. In addition to data gathered by the health subcommittee during the outreach phase of this project, the Working Group survey of UC CIOs and CTOs revealed a number of AI-enabled tools already in use in the health domain. Numerous campuses also intend to adopt AI-enabled tools in the near-term.[39] A few common categories of use cases are:

- **Clinical Diagnosis and Treatment**: AI offers immense potential to improve care delivery, including applications in diagnosis, surgery, personalized medicine and treatment. Indeed in June 2019, UCH and UC hosted the first UC-wide AI in Biomedicine conference, with more than 500 registered UC attendees.[40] The health subcommittee examined one such project below, the use of an AI screening tool for medical imaging.

- **Business Administration**: AI solutions are already being used to improve many aspects of hospital administration, ranging from tasks around scheduling, billing, and payment to fraud reduction and enhancing cybersecurity. For example, the UC Berkeley Optometry Clinic uses AI for the Digital Telehealth appointment booking system for patients. Chatbots are also integrated into UC Berkeley University Health Services. For its second use case, the health subcommittee below describes one of these solutions—an AI-enabled tool to improve the scheduling of appointments used by UC San Diego Health.

- **Population Health Improvement and Chronic Disease Management**: AI-enabled tools can support population health programs in ways ranging from automated screening, to patient self-management tools, to predictive population risk stratification tools. A machine learning risk model developed by UCLA Health serves as the third use case discussed below.

---

[37] "University of California Health Fact Sheet," *University of California Health*, June 4, 2021, https://www.ucop.edu/uc-health/_files/uchealth-at-a-glance.pdf.

[38] Ibid. at 1-3.

[39] An excellent review of the diverse uses of AI in health care can be found in the 2020 National Academy of Medicine Special Report, Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril, Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, editors, National Academy of Medicine, *available at* https://nam.edu/artificial-intelligence-special-publication/.

[40] "A UC-Wide Conference on AI in Biomedicine," *Precision Medicine at UCSF*, accessed July 29, 2021, https://precisionmedicine.ucsf.edu/events/uc-wide-conference-ai-biomedicine

The opportunities described above highlight the potential of AI-enabled tools in the health domain, but there are risks that must be considered and addressed in order for UC to reap the benefits of these tools. In this section, we provide a more detailed look at how the UC Responsible AI Principles can be operationalized in three use cases: (1) AI in medical imaging, (2) scheduling office visits, and (3) UCLA Health Population Risk Model.

AI in Medical Imaging

UC San Francisco researchers, together with GE Healthcare, developed an AI screening tool that works with a portable X-ray machine to screen patients for collapsed lung, or pneumothorax.[41] Pneumothorax affects tens of thousands of Americans each year and occurs when air leaks into the chest cavity causing compression and collapse of the lungs. The screening tool, known as Critical Care Suite, analyzes images of chest X-rays within seconds, flagging any that depict the signs of pneumothorax and sending them directly to a radiologist for review. The process significantly reduces the wait time for interpretation of high-stakes cases, which can reduce suffering and even make the difference between life and death.[42] Critical Care Suite became the first FDA-cleared on-device, point-of-care use of AI in 2019, enabling it to be used in real time in the clinical environment.

To support the development of the tool, UCSF's Center for Digital Health Innovation (CDHI) created a clinical image repository while observing privacy and security principles front of mind, including de-identification and annotation tools and pipelines.[43] CDHI trained the algorithm on thousands of de-identified X-rays that were classified as having or not having pneumothorax. The model was validated using thousands more X-rays, and then tested against additional X-rays from different parts of the world to ensure consistency across diverse patients and care settings. Critical Care Suite is also embedded on-device with the intention of supporting faster and more secure AI analysis, with less reliance on high-bandwidth connectivity.

CDHI worked with GE to expand the Critical Care Suite.[44] For example, Critical Care Suite is being used to help clinicians assess Endotracheal Tube (ETT) placement for intubated patients, including critical COVID-19 patients. The tool analyzes chest X-ray images and precisely measures the ETT position, informing clinicians whether the tube was placed correctly or if it requires further attention. Improper placement of the ETT can cause pneumothorax among other life-threatening complications.

Some early adopters have shared their appreciation for the AI-enabled tool. Amit Gupta, MD, Modality Director, Diagnostic Radiography at University Hospitals Cleveland Medical Center stated, "In several COVID-19 patient cases, the pneumothorax AI algorithm has proved prophetic. It has accurately identified collapsed lungs in intubated COVID-19 patients, flagging the image to radiologists and radiology residents, and enabling expedited patient treatment. Altogether, this

[41] Nina Bai, "Artificial Intelligence That Reads Chest X-Rays Is Approved by FDA," *UCSF Research*, https://www.ucsf.edu/news/2019/09/415406/artificial-intelligence-reads-chest-x-rays-approved-fda.

[42] "FDA Clears GE Healthcare's Critical Care Suite Chest X-ray AI," *Imaging Technology News*, September 12, 2019, https://www.itnonline.com/content/fda-clears-ge-healthcares-critical-care-suite-chest-x-ray-ai.

[43] "The Critical Care Suite: A Case Study in Collaborative Development," *Center for Digital Health Innovation at UCSF*, January 27, 2021, https://www.centerfordigitalhealthinnovation.org/posts/the-critical-care-suite-a-case-study-in-collaborative-development.

[44] "GE Healthcare Announces First X-ray AI to Help Assess Endotracheal Tube Placement for COVID-19 Patients," *Business Wire*, November 23, 2020, https://www.businesswire.com/news/home/20201123005895/en/GE-Healthcare-Announces-First-X-ray-AI-to-Help-Assess-Endotracheal-Tube-Placement-for-COVID-19-Patients.

technology is a game changer, helping us operate more efficiently as a practice, without compromising diagnostic precision."[45]

Principles Implicated & Sub-Recommendations:

- **Transparency:** In designing the Critical Care Suite, the team at CDHI worked to reverse engineer the AI findings to better understand how the algorithms were performing. This included understanding the failure cases to allow iterative improvements to the prototypes. The final product also produces a confidence metric so that end users can conclude how well the AI is performing. Transparency in how algorithms are developed is imperative for gaining user trust and confidence in AI results.

- **Fairness/non-discrimination:** The fidelity of the data utilized to develop AI algorithms remains an important consideration. Data that contains bias can produce bias in the AI algorithms. Robust analysis of data set limitations is necessary to understand potential risks for discrimination in use of algorithms. As such, the Critical Care Suite was tested across a wide spectrum of patient populations worldwide.

- **Shared benefit and prosperity:** The use of AI for medical imaging has the potential to help provide faster and higher-quality care to more patients. Critical Care Suite is specifically designed to be relatively easy for hospitals to adopt because it works with a mobile X-ray machine and does not require new, expensive infrastructure. It can make a big difference in resource-limited care settings that do not have a radiologist on-call at all times. Nonetheless, clinical use of any AI-enabled tool requires staff training and regular monitoring to ensure continued efficacy and benefit for all.

- **Privacy and security:** Privacy and security are critical to the success of AI in medical imaging given the need to train on real patients' medical images, which include highly sensitive information. De-identification is a necessary step, but still does not guarantee that personal information remains private. In addition, de-identification is not always feasible. As hospitals face an increasing number of cyberattacks including ransomware, the security of patient data and the AI systems must also be a primary consideration.

- **Accuracy, reliability, and safety:** Accuracy, reliability, and safety are critical for AI in medical imaging because these tools have influence over life and death decisions. The AI for detecting pneumothorax achieved an accuracy rate of greater than 96% in its validation phase. The tool, which was reviewed and approved by the FDA, also includes an automated AI quality check feature that detects acquisition errors and alerts technologists to review and make corrections as needed before an image is sent to a radiologist. External validation is a key aspect of ensuring that AI technology performs as intended in clinical environments. The tool should continue to be monitored throughout its use in clinical environments to regularly assess the accuracy, reliability, and safety of its use for all patients.

[45] "University Hospitals to begin using artificial intelligence to intubate COVID-19 patients," *WKYC Studios*, May 11, 2021,https://www.wkyc.com/article/news/health/university-hospitals-to-use-artificial-intelligence-to-intubate-covid-19-patients/95-7aea1c56-68c3-490b-8baa-9114a6bd106f.

Scheduling Office Visits

In hospital systems, appointment no-shows can be disruptive and costly. Preemptively identifying appointments that will be canceled can improve patient access to care and provide valuable cost savings to operations. AI-based scheduling tools can help to optimize scheduling by predicting potential no-shows, thereby improving provider efficiencies and patient satisfaction. Such tools may not, however, identify or address the root cause of no-shows, thereby potentially perpetuating inequities in access to healthcare. When procuring or using predictive modeling for operational decision support, it is important to think about downstream ethical, legal, and social implications prospectively, and mitigate risks appropriately.

The UC San Diego Health Enterprise AI committee discussed this particular use case in 2020. Specifically, an algorithm available via UCSD's electronic health record system would facilitate identification of potential no-shows and manage that possibility by allowing for double-booking that appointment. While this may, at face value, appear to be an innovative solution, in practice, the model might not factor in social determinants that influence a patient's ability to make their appointment. This could result in low-resourced and marginalized populations being double-booked more often than others. When testing the model on existing data, the results confirmed this outcome. The Enterprise AI committee met to discuss how best to mitigate possible harms if the system were to be implemented.

Principles Implicated & Sub-Recommendations:

- **Appropriateness*:*** The potential benefits to operations must be weighed against the possible risks of harm to and needs of the patients served by the health system. The possibility of greater efficiencies from a business model may be compelling. However, if those who are double-booked end up having to wait longer for their appointment, it could lead to unanticipated consequences.

- **Human Values*:*** The ideals of human values, in particular treating every patient equally, come into play when social determinants of health (SDoH) are influencing the ability of a patient to miss or make an appointment. In addition to incorporating SDoH factors into the model, healthcare providers and staff also have an opportunity to identify whether those missing appointments need services to assist in keeping their scheduled appointments.

- **Transparency**: The UC healthcare providers and patients should be involved in discussing potential ramifications of AI-enabled appointment management tools.

- **Fairness and Non-Discrimination*:*** Disclosure of how the algorithm is trained as well as accountability must be visible. If the model is trained on unrepresentative data, the results will likely cause harm. The model should be tested using retrospective patient data to determine the possible effect on patients. Provided bias is managed, the function of the algorithm should be checked periodically to assess benefit vs risks.

- **Shared Benefit and Prosperity*:*** AI-enabled scheduling tools should incorporate insights from healthcare staff and patients on possible social, reputational, and economic harms and determine if those possible harms are able to be mitigated and offset by possible benefits to patient care.

- **Privacy and Security**: Privacy and security are important to the extent that exposure of patient information leads to discrimination and stigmatization and should be at the forefront of decision-making.

UCLA Health Population Risk Model

In late 2018, UCLA Health began development of a machine learning model to predict the risk of hospitalization and/or emergency department (ED) visits over the next 12 months in primary care patients. The goal of developing this risk model was to help patients avoid unnecessary ED visits and hospitalization by using risk scores to identify—and then proactively conduct outreach to—these at-risk patients to coordinate their care, encourage self-management, address social determinants, and ensure completion of physician care plans. The design and implementation of the model involved broad collaboration and vetting across UCLA Health, incorporating input from executive leadership, health informatics and analytics, clinicians, population health experts, legal and compliance, and ambulatory care coordination.

In undertaking this project, UCLA Health set out to construct an outcome that would be a good proxy for unmet patient health needs and focused on three criteria: that it be clinically significant, that it be preventable, and that there be sufficient lead time for intervention. After deciding on the risk of hospitalization and/or ED visits over the next 12 months as its desired outcome, the team developed the model utilizing numerous data elements from categories such as demographics, past utilization, health conditions, and other clinical data. These elements were derived from EHR data, administrative claims data, and the Area Deprivation Index.[46] The Area Deprivation Index is a census derived index that can be used as a proxy for SDoH. Since the privacy and security of the data were top priorities, UCLA Health developed the machine learning algorithm in a secure UCLA Health environment maintained by UCLA Health's Office of Health Information and Analytics (OHIA). The team fed the algorithm with data on its 400,000 primary care patients and it returned approximately 6,000 patients at risk of hospitalization or emergency room visits over the next 12 months. Patient lists are generated quarterly and empaneled to a team of nurses, social workers, care coordinators, administrative staff, and physicians who work proactively with patients to coordinate their care and address social determinants. Recognizing that the model does not identify all at-risk patients, the team also provides a process whereby physicians have been able to utilize their own clinical judgment to identify additional high-risk patients. This model of care termed the "Proactive Care Model" has since been implemented in 50 UCLA Health primary care practices across Southern California and a preliminary review of the data has shown a trend in (and potentially statistically significant) reduction in hospitalization and ED visits since implementation.

Principles Implicated & Sub-Recommendations:

- **Fairness/non-discrimination.** The UCLA Health team recognized the potential for the model to incorporate bias and systemically and unfairly advantage certain groups. For example, a 2019 study by Obermeyer et al. found evidence of racial bias in a predictive risk model deployed nationwide to identify and help patients with complex health needs.[47] This

[46] "Neighborhood Atlas," Department of Medicine, School of Medicine and Public Health, University of Wisconsin, accessed Aug. 15, 2021, https://www.neighborhoodatlas.medicine.wisc.edu/.

[47] Ziad Obermeyer et al., " Dissecting racial bias in an algorithm used to manage the health of populations," *Science* 366 (2019): 447-453, https://doi.org/10.1126/science.aax2342.

model used patient expenditure as the predicted outcome and assigned risk scores to patients, with those above a certain threshold targeted for intervention. Since less money is spent on Black patients who have the same level of need, the algorithm falsely concluded that Black patients are less in need of intervention. The researchers found that reformulating the algorithm so that it no longer used costs as a proxy for needs eliminated the racial bias in predicting who needs extra care. Consistent with these findings, the UCLA Health team decided to use ED visits and admissions as a proxy for unmet health needs and incorporated in its model validation process an effort to detect bias. Model validation confirmed that the algorithm's outcome greatly reduced or perhaps even eliminated the bias documented by Obermeyer. Despite the above success in reducing biases, the subcommittee notes that removing the cost from the model can be insufficient to guarantee a fair treatment. This is often referred to as the *fairness through unawareness* argument. Depending on different applications, it may be possible that there are other factors that would correlate with the cost. Therefore, the model will inherit the bias from other factors, even though the cost parameter is removed from consideration. Models should be subjected to periodic review to mitigate unintended bias and ensure the reliability of their use.

- **Privacy and Security**: The use of large datasets to develop and train machine learning algorithms always presents risks to the privacy and security of the data that teams must prioritize and address. Models should be developed in secure environments that do not allow for the downloading of data.

- **Transparency**: UCH should be transparent about their development and use of models and incorporate input from a diverse set of stakeholders.

Recommendations on Implementation of UC AI Principles in Health

This section examined the benefits and risks arising from current and potential uses of AI in healthcare and analyzed the ways in which current uses of AI-enabled tools for UC health data and services implicate the UC Responsible AI Principles. The findings highlighted how UC can use AI-enabled tools to help support diagnostic imaging, identify workflow optimization, and improve population health. At the same time, UC needs strategies in place to protect itself and the people it serves from unique and exacerbated risks associated with the use of AI-enabled tools to the privacy and security of health data, equitable health treatment, and the protection of people's safety and rights.

Based on the above considerations, the health subcommittee recommends that UC engage in the following next steps to implement the UC Responsible AI Principles into the health domain:

1. Advance the development and implementation of an AI risk and impact assessment framework that stakeholders throughout UCH can use to inform decisions about the design, procurement, and use of AI.

2. Encourage documentation practices for AI systems to facilitate assessment, transparency, auditing, and oversight.

3. Develop and incorporate standardized, reproducible, and meaningful processes for human review, testing, evaluation, and monitoring at appropriate checkpoints throughout the AI lifecycle to ensure that deployed AI systems support ideals of human values and consistency with the UC mission.

4. Encourage representation, engagement, and feedback from the UC community and relevant external stakeholders at all phases of AI exploration and adoption.

5. Provide training and educational programs and ensure adequate allocation of funds, both at UCH and at the individual health campuses, to further development and local implementation of the best practices and standard operating procedures suggested in this report.

6. Develop guidelines for AI procurement in the health domain, including guidance on monitoring and stewardship.

7. Conduct further research on the potential and experienced benefits and risks of AI throughout UCH as well as compliance with relevant laws, regulations, and standards.

Each of these recommendations is explored in greater detail below.

1. AI Risk and Impact Assessment Framework

The scale and scope of risks and impacts associated with UC Health AI will vary based upon the type of AI technologies employed, how and for what they are used, and the broader context of their implementation. Indeed, assessment of relative risk and benefits is foundational to medical decision-making, including the use of diagnostic testing and treatments. For example, comparatively simple linear regression models used to help automate back-end hospital processes may be less risky and consequential than using deep learning to help detect cancer, support surgeries, or stratify patients for risk management purposes. We encourage UC to consider developing a framework to assist software developers, purchasers, and users of AI technologies to understand that greater oversight will be needed in cases where people's health, safety, and rights are directly at stake.

This recommendation will help to operationalize the first principle of appropriateness from the UC Responsible AI Principles, which states, "the potential benefits and risks and the needs and priorities of those affected should be carefully evaluated to determine whether AI should be applied or prohibited." Robust risk assessment will also help to operationalize other AI Princples, including accuracy, reliability, and safety, fairness and non-discrimination, and privacy and security. Specific risk mitigation measures should be associated with each level of risk and impact, with more stringent accountability and review processes for AI-enabled tools that present higher levels of risk to patients, or to the organization. Recognizing that applications are unlikely to be zero-risk, multidisciplinary teams will need to determine acceptable risk and effective mitigation measures depending on the context. Applications deemed to have unacceptable risks to human safety or rights should not be pursued further. Above all, as discussed above, use of AI in healthcare is intended to complement human intelligence, not replace it.

2. AI Documentation

One of the UC Responsible AI Principles is transparency, which stipulates that the UC community should be made aware when UC is proposing and/or using AI-enabled tools. Key elements for UC health to keep in mind include records of what AI technologies are used, where, and for what purposes. Additionally, information about the design of the AI systems (such as what training data were used and why) and any weaknesses or risks associated with its uses should be included to support additional UC Responsible AI principles. Documentation regarding the design of the AI systems also includes whether and how a third party might be using UC Health data for the purposes of training or enabling AI systems, as well as how risks related to such use are being mitigated. To the extent possible, documentation practices should be standardized to enable efficient and comprehensive evaluations of common use cases in health AI both within and between UC Health campuses. For example, one idea that has been raised to achieve standardization and increase transparency is to use a "model facts label" for implemented models which would be publicly available.[48] Robust documentation will promote transparency, enable independent audits, and also play a critical role in the responsible design, use, and monitoring of AI in the UCH domain.

3. Human Review and Monitoring

Cutting-edge AI systems still suffer from problems related to bias, errors, reliability, and security vulnerabilities; responsible use requires ongoing oversight, including processes for testing and monitoring the impact of these systems in practice. We recommend that UCH implement processes to incorporate meaningful human review at appropriate checkpoints throughout the AI lifecycle to mitigate real-world harms caused by known and unforeseen technical limitations, misuse, or complications and consequences that emerge in the context of real-world applications. AI technologies may well become a valuable aid for clinicians and others throughout the health domain, but they are unlikely to fully automate current jobs or replace human decision-making in most circumstances, and it is essential that there be processes in place to ensure patient safety when these technologies are in use. Monitoring performance of AI-enabled tools—including false positives and false negatives against baselines—should also include assessing impact on different communities to ensure the principles of fairness and non-discrimination and shared benefit and prosperity are respected.

4. Representation and Engagement

The field of AI is evolving rapidly, and the desires and concerns of the UC community and relevant external stakeholders—including patients and healthcare providers and staff—may similarly shift over time. UC should encourage relevant and affected communities to provide input and feedback on the use of AI throughout the health domain during the design phase of significant projects and following deployment. To support the operationalization of the principle of transparency, this recommendation includes promoting communication techniques and interfaces that help make AI systems understandable to diverse patient communities and providing ways to challenge outcomes and address any harms caused.

5. Training and Educational Programs

The subcommittee found that many UCH locations have already begun to implement AI-enabled tools throughout health-related processes and services. We understand that there is interest in AI

[48] *See*, *e.g*, https://www.nature.com/articles/s41746-020-0253-3.

training and educational programs to help UC staff make thoughtful decisions about where and how AI-enabled tools may be appropriate and beneficial, and how to appropriately assess ethical implications and mitigate associated risks and impacts. We encourage UC to support the provision of relevant training and support across these different topics. As one of the largest medical training environments in the US, UC's actions in this regard will have a significant long-term impact on clinical practice. Investment in this effort, including adequate funding to support the expertise required for thorough evaluation and documentation of AI-enabled tools—both before and after adoption—is critical to ensuring that best practices are implemented.

6. Procurement Best Practices

Our subcommittee recommends that UC develop best practices regarding the procurement of AI-enabled tools, in furtherance of many of the UC Responsible AI Principles, including transparency, appropriateness, shared benefit, fairness, privacy and security, accuracy, reliability and safety. As the UC Executive Director of Strategic Sourcing raised during his presentation to the full Working Group, these best practices might include strategies for AI adoption; risk assessment; setting boundaries around unacceptable uses of AI; and governance structures, including those that address AI embedded in other products under consideration.[49] In determining best practices, factors such as FDA review of products that fall within its purview may provide a useful means for evaluating the accuracy, reliability, and safety of these products.

7. Further Research

Our subcommittee recommends that UC conduct further research on the potential and experienced benefits and risks of AI throughout UCH as well as relevant laws, regulations, and interrelated technical, ethical, and societal factors associated with the use of AI in the health domain. Additional topics meriting further consideration may include: effective methods of monitoring AI to promote adherence with the AI Ethical Principles; potential over-reliance on digital platforms; the role of a profit motive in the development of UC AI systems; and how to communicate with and incorporate feedback from patients regarding the use of AI tools by administrators and clinicians. This recommendation is intended to support the implementation of the previous recommendations.

[49] Justin Sullivan, Executive Director, Strategic Sourcing Centers of Excellence, UCOP, presented to the UC Presidential Working Group on AI on June 16, 2021.

Introduction

The UC AI Working Group formed a human resources (HR) subcommittee due to the increasing use of AI-enabled tools in HR and the benefits and risks they pose. The subcommittee evaluated the current use of, prospective opportunities for, and hazards of using AI software and related data-driven programs across five employee life-cycle roles (i.e., applicants, prospective employees, new employees, current employees, and retired employees) and nine stages (i.e., recruitment, onboarding, learning and professional development, performance assessment, mentoring, retention, career development and advancement, separation, and post-separation). The subcommittee also debated what constitutes "fairness" for AI-enabled systems used to support HR activities within UC and outlined ways to ensure that the proposed UC Responsible AI Principles are respected by HR systems.

The motivation behind forming this subcommittee was an anticipated increase in the use of automation in HR processes within UC. Both AI and robotic process automation (RPA) are increasingly being integrated into HR workflows. While these two concepts are similar in that they both automate tasks traditionally performed by humans, RPA relies on inputs and logics designed by humans and thus cannot be said to deploy its own "intelligence," while AI is a broader concept that can include machine learning processes, which rely on algorithms that empower a digital system to "learn," "reason," and "self-correct," and thus adapt to and address issues that were not explicitly coded for.[50] While RPA is especially on the rise in the HR sector, the use of AI is expected to increase in the near future. Thus, we reference opportunities and concerns related to both throughout this report.

With UC employing more than 185,000 people and serving as one of the largest employers in the state, the stakes are high to responsibly integrate AI into HR.[51] Recent attempts to legislate the deployment of AI across a range of processes, including HR—both to spur innovation and to minimize the risk of potential misuses and unintended consequences—suggest that now is an ideal time to identify trends and potential use cases, and evaluate how UC might align its planning with

[50] "AI and RPA: What's the Difference and which is Best for Your Organization?," *NICE*, accessed May 17, 2021, https://www.nice.com/rpa/rpa-guide/rpa-ai-and-rpa-whats-the-difference-and-which-is-best-for-your-organization/#:~:text=While%20RPA%20is%20used%20to,and%20develops%20its%20own%20logic

[51] "Working at UC," *University of California Website*, accessed June 28, 2021, https://www.universityofcalifornia.edu/uc-system/working-uc.

emerging standards and principles, including those proposed by the UC AI Working Group, in the adoption of AI-enabled tools for HR purposes.[52]

AI is "changing how work gets done… by making operations more efficient, supporting better decision-making, and freeing up workers from repetitive tasks."[53] This includes the potential to make processes and outcomes more effective and equitable. More specifically, AI-enabled tools can help "automate, assist, or augment the heavy lifting of transaction, research, and analysis work" that HR requires.[54] HR teams integrate AI for a number of reasons, which include managing large volumes of applications; meeting applicant demand; sourcing hires; facilitating effective hiring and evaluation process; reducing costs; and supporting equity, diversity, and inclusion (EDI) initiatives.[55] For this report, the HR subcommittee especially focused on EDI-related impacts, noting that EDI is not just about diversity, but also having a sense of belonging and personal connection to UC and its disparate locations, which could be undermined by automated processes that feel depersonalized.

In the HR context, AI is taking the following forms: speech recognition and natural language processing (e.g., for sentiment analysis); computer vision and optical character recognition (e.g., scanning resumes for job suitability); rule-based systems and predictive analytics (e.g., predicting team turnover or retention); machine learning and deep learning (e.g., scanning recruitment sites for valuable data for candidate searches); and RPA (e.g., shifting routine and repetitive tasks from humans to machines).[56]

Use Cases for AI in HR

The HR subcommittee organized its analysis around eight HR "stages" identified as particularly critical to the employee lifecycle: (1) pre-recruitment; (2) recruitment, (3) onboarding, (4) employee wellbeing and engagement, (5) learning and development, (6) career development and advancement, (7) separation, and (8) post-separation.[57] Below, we provide example use cases of how AI is being used or may be used within UC. While some issues arise at multiple points throughout the employment lifecycle (e.g., salary adjustments, or comfort with automated interactive processes), we mention recurring issues where especially salient.

---

[52] "Legislation Related to Artificial Intelligence: 2021 Legislation," *National Conference of State Legislatures*, https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to -artificial-intelligence. aspx.

[53] "Better decision making" is understood for purposes of this memo to mean decision-making informed by the comprehensive analysis of the best available data, leveraging very large datasets in tandem to identify patterns and generate insights, often in an automated fashion that enhances speed and efficiency of such decision making.; Susanne Hupfer, "Talent and workforce effects in the age of AI: Insights from Deloitte's State of AI in the Enterprise, 2nd Edition Survey," *Deloitte*, March 3, 2020, https://www2.deloitte.com/us/en/insights/focus/cognitive- technologies/ai- adoption-in-the-workforce.html.

[54] Chris Havrilla and Charu Ratnu, "5 Categories of AI Commonly Used in HR," *Deloitte*, November 12, 2019.

[55] Brandie Nonnecke, "The New HR: Employing Equitable AI," accessed May 17, 2021, https://docs.google.com/presentation/d/17G08p7sfKEdUFJotg5MSQvSAcwFKJQGns_eR9Aez7V8/edit#slide=id.g80ca03e1fe_0_215.

[56] Ibid. Major issues that arise from an HR perspective include whether UC should reskill those being replaced by AI; whether this would be seen as "contracting out," which could be a problem for positions covered by collective bargaining; and the quality of the human experience (for example, when an employee engages with a robot on personnel issues and feels dehumanized because they don't get the information they need, does there need to be an option for human engagement?).

[57] There are multiple ways to categorize HR processes. We have chosen to organize chronologically around the worker experience. However, organizations such as Deloitte have instead bucketed the potential functions of AI around the employer experience: Talent Planning and Acquisition, Workforce Movements and Data Management Activities, Growth and Development Activities, HR Operations and Support Activities, Worker Relations, and Total Rewards (such as payments and incentives, processing leave requests, etc.).

The use cases illustrate the opportunities and challenges of applying AI across eight HR "stages":

- **Pre-recruitment:** Examples of AI in pre-recruitment include its use to develop job descriptions that may attract a broader demographic diversity of applicants and to identify and assess qualifications that may be especially helpful in particular jobs.

- **Recruitment:** AI is already commonly used during recruitment in the marketplace outside UC. Relevant practices include the automation of reviewing resumes to mitigate discrimination and bias and reduce time required to identify qualified candidates, as well as conducting candidate assessments and setting appropriate salaries, with an eye toward equity.[58]

- **Onboarding:** AI may be used to document new hire data by automating the transactional processes and determining which onboarding tasks need to be presented to the new employee. AI has also been used to measure the effectiveness of onboarding processes, and to customize the experience for each employee based on attributes of their new job or as individuals.[59]

- **Employee Wellbeing and Engagement:** AI holds great potential to streamline mental wellness support and connection with benefits providers. For some employees, using AI to triage support resources may help to address their needs more quickly and confidentially. AI is being used to track employee-identified wellness goals, progress, and provide nudges to encourage healthy behaviors that may qualify for firm-provided health incentives. In addition, some organizations provide and encourage employees to adopt wearable technology that can track and process various health, fitness, and stress data, including users' heart rate, daily exercise, gym use, and more.

- **Learning and Development:** AI is especially relevant to learning and development in three ways. First, as AI-based tools and processes are integrated into existing workflows, what is or should be the obligation of UC to train employees to use those AI-enabled tools, or to reskill employees if or when they are replaced by AI? Second, using AI to automate the process of identifying candidates for additional training, as well as tracking the effectiveness and impact of training programs over time. Third, using AI to improve learners' experience by curating and recommending content based on their skills and interests, previous learning activities, and learners with similar profiles, currently seen in digital learning platforms such as Skillsoft Percipio and LinkedIn Learning.[60]

- **Career Development and Advancement:** AI may identify who is ready for additional training or advancement opportunities, as well as what positions employees may be qualified to apply for given knowledge, skills, abilities, experience, and training. AI might also be used to ensure equity in salary increases across job classifications and job functions, and

---

[58] Genevieve Smith, Associate Director - Center for Equity, Gender & Leadership, University of California Berkeley, Haas School of Business, Interview by Mark Cianca and Alexa Koenig, February 12,. 2021.; For incredibly helpful and timely resources addressing issues related to equity and AI, including those that relate to HR processes, see Berkeley Haas Center for Equity, Gender, and Leadership, "Mitigating Bias in Artificial Intelligence," https://haas.berkeley.edu/equity/industry/playbooks/mitigating-bias-in-ai/.

[59] "AI in HR: Assessing Automation Opportunities," *Deloitte*, 2019.

[60] "Meet Skillsoft Principio," *Skillsoft*, accessed July 15, 2021, https://www.skillsoft.com/meet-skillsoft- percipio.; "LinkedIn Learning," *LinkedIn*, accessed July 15, 2021, https://www.linkedin.com/learning/me.

to help identify and remedy any imbalances based on areas of historic discrimination. However, application of AI in this domain simultaneously poses significant risks to equity and inclusion if deployed without appropriate analysis and oversight.

- **Separation:** AI is expected to have several uses and effects related to separation. We focus on two: (1) AI as a tool that supports the separation process and (2) AI as something that replaces human functions, forcing people out of their jobs as their skills become obsolete. With regard to the first, AI could effectively be used to automate the removal of access and other security steps needed upon separation.[61]

- **Post-Separation:** AI considerations may extend into the post-separation phase of former employees' relationship with UC. For example, AI might be used to support benefits and pension planning and other types of financial analysis to help former UC employees prepare for and transition into a secure retirement. UC should ensure that both current and former employees have a degree of comfort with the automation of services, including voice recognition and other processes. Notably, a lack of familiarity and comfort with such AI interventions may—if not properly implemented—compound the trauma of loss and other stressors affiliated with major life transitions.

In this section, we provide a more detailed look at how the UC Responsible AI Principles can be operationalized in three use cases: (1) recruitment, (2) workflows, and (3) compensation management and pay equity.

Recruitment

Recruiters are increasingly using AI to identify potential applicants who satisfy minimum qualifications for open positions. For example, using AI to comb sites like LinkedIn to identify individuals who meet required minimum job qualifications defined in the job description, who may have additional desirable credentials or relevant work histories, and who can then be targeted for recruitment. Such uses are already beginning to occur within UC.[62]

We assessed the use of AI for recruitment as having enormous positive potential but also enormous risk. While the use of AI in recruitment could help mitigate human bias (e.g., by helping to identify traditionally underrepresented individuals with applicable skill sets), it can also magnify bias (e.g., by privileging potential candidates from elite universities, whose status may reflect long-standing biases in opportunity).[63]

Most studies that have purported to assess the positive and negative impacts of AI-based recruitment processes on underrepresented individuals have been small and not statistically valid, meaning there's a lack of trustworthy data regarding those effects. However, the potential use of such technologies to comb huge datasets to find individuals who would likely thrive in the position and provide critical diversification of staff and faculty is worth exploring. Automated identification of common variables among successful past employees may help to counter human biases regarding

[61] "AI in HR: Assessing Automation Opportunities," *Deloitte*, 2019.

[62] Genevieve Smith, Associate Director, Center for Equity, Gender & Leadership, University of California Berkeley, Haas School of Business, Interview by Mark Cianca and Alexa Koenig, February 12, 2021.

[63] Use of AI to mitigate human bias would need to be implemented in a way that did not run afoul of the UC's legal responsibilities including under Proposition 209, the Equal Protection Clause and Title VII.

what makes a "successful" potential employee, while algorithms can be adjusted to ensure that traditionally underrepresented individuals are prioritized (or at a minimum not disadvantaged) during automated recruitment processes.

On the negative side, potential candidates can also use AI to game the recruitment system. This may affect the utility of some forms of transparency around the algorithms and datasets that underlie the AI system and how UC might deploy it.[64] In addition, the use of job bots to automate management of the candidate experience should be carefully reviewed to prevent any unintended lack of transparency and to avoid any violation of "human values"—one of the UC Responsible AI Principles—such as utilizing a bot that is not in sync with an individual's social and cultural values.

> **"Risks might be addressed through an ongoing review process that integrates experts who study AI in HR."**

Another area of sensitivity and concern is privacy. Some people may avoid joining online platforms that AI systems rely on for candidate recruitment, like Facebook, Twitter, TikTok, or LinkedIn, because of privacy concerns. Others may join social media sites but refrain from sharing personal information, given privacy concerns and gender, racial, and other characteristics with potential for discrimination and bias online. Additionally, it's important to consider how use of AI to comb social media data affects the quality of the search process, disproportionately excluding those who choose not to share private information and those without reliable access to the Internet. If these individuals might miss out on employment opportunities then such a negative outcome from the standpoint of equity would argue against UC's use of such a tool.[65]

If AI is used to develop a diverse candidate pool, and the AI is tuned to search for data commonly used for hiring processes, the University should periodically review and audit the impact and outcomes, and tune the AI to adapt to UC legal and compliance requirements, norms and trends in privacy, and other legal and social justice considerations. Mitigating risks might be addressed through an ongoing review process that integrates experts who study AI in HR, including UC HR talent acquisition leaders who can provide data and feedback regarding current uses, benefits, and challenges. In addition, contracts with third-party providers for AI-enabled HR services could include requirements for periodic review and adjustment.

Another important consideration is *how* to use AI appropriately, whether for analysis or for routine tasks. For example, AI might be used to anonymize or de-identify resumes and other applicant materials to minimize human discrimination and bias in analysis. Generally, the subcommittee felt UC should avoid using AI to make judgment calls about candidates in order to avoid the "eugenics of automation" (e.g., using physical characteristics to discern elements like "dependability") and instead prioritize its use for more transactional tasks (such as anonymizing resumes). The subcommittee also noted that in some cases, de-identification may actually work against diversity, equity, and inclusion, spotlighting this as something that should be carefully considered when debating current and future uses of AI.

---

[64] Of course, disclosure may be compelled—or attempted—through Freedom of Information Act requests.

[65] For background information on the UC's use of social media profiles in recruitment see University of California Office of the President, "Recruitment Through the Use of Social Networks," April 21, 2020, https://policy.ucop.edu/doc/4000582/SocialMediaRecruiting.

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** Every current and prospective use of AI for recruitment should be evaluated for its appropriateness for meeting UC-wide objectives and legal compliance requirements—one of which should be adherence to the UC Responsible AI Principles. For example, use of AI should support more "effective" and "equitable" recruitment of individuals who are qualified and well-suited for open positions. Use of AI should help UC better understand what qualities foster success in open positions rather than relying on historic assumptions, which may be more tightly correlated with categories of systemic implicit bias/discrimination, including race/ethnicity, gender, and class, rather than actual qualifications and capability.

- **Human Values:** Automation of interactions with candidates and prospective candidates—whether through AI or RPA—should not infringe on human agency and dignity, or civil and human rights by depersonalizing those interactions. The use of AI should be respectful and protective of privacy and other legal and social concerns.

- **Transparency:** HR units should be transparent about when and how they are using AI for recruitment and the potential for candidates to adapt their behavior accordingly or ensure their qualifications are accurately presented. Transparency should be constrained in ways that minimize the risk that more technologically sophisticated individuals will use that transparency to "game" recruitment processes by artificially or unfairly increasing their visibility to algorithms.

- **Fairness/Non-discrimination:** The use of AI should be calibrated to minimize discrimination and bias, and maximize the diversity of applicant pools. The use of such AI should be "fair" to those potential candidates who may have had less exposure to AI systems or who may not participate online, including on social media or other platforms where data are collected and analyzed by AI HR tools.

- **Shared Benefit and Prosperity:** The use of AI in recruitment should benefit both the university and potential applicants, especially those who may not traditionally have been identified as top candidates through recruitment processes but who could provide critical skills and enrich diversity.

- **Privacy and Security:** UC must consider whether the use of AI would force potential candidates to distort their privacy practices, including what they reveal online, to make themselves more visible to machines. Prospective candidates should be alerted to the information being collected about them, the information collected about prospective candidates must be limited to what is relevant to the job description, kept secure, and retained only so long as is necessary to complete the purpose for its collection.

- **Accuracy, Reliability, and Safety:** Underlying datasets and algorithms must be regularly updated to ensure ever-improving accuracy and reliability in the recruitment process. Use of AI should actually do what it claims to do (assembling a diverse and highly qualified applicant pool) and should not be biased or otherwise unfairly distort processes or outcomes.

- **Accountability:** UC must hold itself accountable for the use of AI in recruitment processes, including implementing appropriate response and remedy procedures for problematic use or outcomes.

Workflows

The use of AI can both positively and negatively affect employee satisfaction and workflows. Common uses can include employee performance assessments, as well as identifying employees who may be ready for additional training and potential advancement.

The use of AI and RPA, when integrated into employee workflows, can also dramatically affect job satisfaction. RPA in particular is already being used to reduce repetitive tasks through automation and provide workflow support. Ernst and Young (EY), a global consulting firm, partnered with a large US public higher education system to utilize RPA to automate reconciliation between files on two HR systems, saving the institution approximately 20 hours in manual processing hours per month.[66]

While satisfaction may increase if AI removes tasks that employees dislike, satisfaction may decrease if automation replaces tasks an employee enjoys. Because perceived job security is a critical component of job satisfaction, replacing employees' tasks may be a source of stress if the individual believes core duties are being replaced by a machine. Ultimately, the net value of "outsourcing" jobs to AI is an important issue to consider—from both a cost-savings perspective but also effects on overall job satisfaction.

A critical question for the UC is "what are our ethical obligations to those faculty/staff who are replaced by automation?" And what are the processes in which UC should responsibly engage when that happens? UC's regulatory environment and collective bargaining agreements include specific requirements related to the activities of outsourcing work. Leaders will need to clarify whether the deployment of AI or RPA that displaces a workforce raises any of these issues and partner with unions to address how this disruption might affect collective bargaining agreements.

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** UC must consider the appropriateness of implementing AI or RPA. Implementation for repetitive tasks that machines are better suited to engage in than humans can be beneficial, freeing people to do higher-level reasoning or more "interesting" tasks. However, UC must also consider the effects of AI or RPA on employee satisfaction, recognizing that cost or time savings may not be the only or even best rationale for its deployment.

- **Human Values:** Integration of AI or RPA into the workflow can affect employees' sense of dignity and agency. UC should put in place assessments to ensure employees are benefiting from the integration of these technologies and minimizing harms to worker agency and satisfaction.

- **Transparency:** UC must consider the need for transparency about when it is considering integrating AI or RPA into workflows, or for assessing candidates for learning, development,

[66] Brad Duncan and Kasia Lundy, "How Universities are Using Robotic Process Automation," *EY*, accessed July 20, 2021, https://www.ey.com/en_us/government-public-sector/how-universities-are-using-robotic- process-automation.

and/or advancement, so that employees can provide critical feedback and input on the decision-making process.

- **Fairness/Non-discrimination:** The integration of AI or RPA should not negatively or disproportionately affect the job experiences of employees from underrepresented demographics. If so, UC must take action to alleviate this difference.

- **Shared Benefit and Prosperity:** The integration of AI or RPA should benefit not only the University as a whole but also the individual employees whose positions are affected, for example by providing new opportunities for learning, growth, and development of their abilities and careers, or eliminating less interesting and repetitive tasks.

- **Privacy and Security:** The University must consider whether the integration of AI or RPA raises new privacy or security concerns, for example by centralizing data about employees' readiness for advancement or other potentially sensitive personnel information.

- **Accuracy, Reliability, and Safety:** UC must minimize the risk that AI systems inequitably target certain individuals for learning and development over others.

- **Accountability:** The University must hold itself accountable for the development and use of AI and RPA to ensure the equitable and timely identification of employees for advancement. It must put in place appropriate correction procedures if AI or RPA uses are found to dehumanize or otherwise violate any of the principles touched on above.

Compensation Management & Pay Equity

Pay equity is where "the rubber meets the road" with regard to University rhetoric around treating employees well. Today, employees widely doubt organizations' ability to ensure pay equity.[67] AI—when responsibly deployed—may help not only with the quality of equity in any compensation system (e.g., by detecting pay inequities within classifications and work functions or across demographics), but also perceptions of equity if AI uses are well explained to employees, potential employees, and the broader public.

AI can be leveraged to collect and "price" skills from various sources, understand which roles require certain skills, and apply geographic differentials, a common HR compensation practice.[68] However, the risks of AI-perpetuated bias need to be fully understood and managed to ensure the mitigation of systemic issues inherent in job descriptions, unrepresentative sampling, and reliance on flawed data. For example, if an AI-enabled tool is taught by feeding compensation data from a population that already reflects inequities, the resulting algorithm will likely perpetuate these inequities.

A burning question is whether AI is effectively able to correct past pay discrimination and prevent such discrimination in the future. While research suggests the use of AI rarely eliminates pay differentials, countering human biases in setting salaries seems like a particularly powerful and

[67] Tanya Jansen, "How AI in HR Will Close the Gender Pay Gap," *HR Technologist*, February 12 2019, https://www.hrtechnologist.com/articles/digital-transformation/how-ai-in-hr-will-close-the-gender-pay-gap/.
[68] Joanne Sammer, "Bringing Artificial Intelligence into Pay Decisions," *SHRM*, December 10, 2019, https://www.shrm.org/resourcesandtools/hr-topics/compensation/pages/bringing-artificial-intelligence-into-pay-decisions.aspx.

important potential use of AI—a process that some have described as "more art than science."[69] UC should stay abreast of latest findings about how, when, and under what conditions AI has been found to minimize bias in salary setting, and how it has been deployed without significantly increasing legal risks.[70] An additional requirement is that UC must consider applicable legal compliance issues such as California Assembly Bill 168 (AB 168), which prohibits employers from inquiring into an applicant's salary history, effective January 1, 2018.[71] The legislation is intended to address inequity in pay practices based on gender, race, color, religion, sex, national origin, disability, age, protected veteran status, gender identity, or sexual orientation. Given UC's institutional values and commitment to ensuring equal pay, it has aligned policies and practices with the provisions in this law and must continue to ensure that future systems, processes, and procedures align and comply.

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** Given that AI has yet to prove particularly effective for eliminating pay disparities, UC should carefully scrutinize any prospective use in this area, review any legal and compliance requirements, and pilot test its deployment to evaluate its efficacy.

- **Human Values:** If an AI-enabled tool is superior to a human at assessing pay equity (and allows for evaluation across a range of variables, including classification, work function, and demographic information), such software could have a significant positive impact on employees' sense of dignity (including that the University respects their labor), as well as civil and human rights (e.g., if pay disparities are minimized or eliminated across race/ethnicity, gender, ability, sexual orientation, and more).

- **Transparency:** If AI is deployed in pay-setting processes, UC should be transparent about its use and establish appropriate processes to enable and facilitate human review of potential problems.

- **Fairness/Non-discrimination:** Any software used for pay equity purposes should be carefully assessed with regard to what it is evaluating and the datasets used to ensure that automation is not masking or shifting biases in problematic ways.

- **Shared Benefit and Prosperity:** Ultimately, this use of AI has the potential to result in greater worker satisfaction and productivity in ways that benefit both the University and employees.

- **Privacy and Security:** Given that this use case incorporates highly sensitive demographic and pay history data, evaluators should especially scrutinize the privacy and security risks of deploying AI for pay equity purposes.

---

[69] Ben Eubanks, "Can Artificial Intelligence Solve the Pay Gap Problem?" *upstartHR*, June 4, 2018, https://upstarthr.com/can-artificial-intelligence-solve-the-pay-gap-problem/; Ben Eubanks, *Artificial Intelligence for HR: Use AI to Support and Develop a Successful Workforce* (Kogan Page, 2018).

[70] Michelle Capezza and Bradley Merrill Thompson , "AI-Based Compensation Management and Bias: Can AI Close the Pay Gap?" *National Law Review*, March 11 2021, https://www.natlawreview.com/article/ai-based-compensation-management-and-bias-can-ai-close-pay-gap.

[71] Personal Rights: Automated Decision Systems, A.B. 2269, 2019-2020 Reg. Sess. (Cal. 2020), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB2269.

- **Accuracy, Reliability and Safety:** The use of AI for pay equity should be regularly audited and checked against human assessments to ensure accuracy, reliability, and safety and that its use isn't creating new frustrations and concerns among employees and their supervisors.

- **Accountability:** If the University were to deploy AI to drive pay equity, it must develop a plan to remedy any misuse or problematic outcomes that occur.

Recommendations on Implementation of UC AI Principles in HR

This section examined the benefits and risks arising from current and potential uses of AI in HR. The use cases highlight ways in which UC can utilize AI-enabled tools to support greater efficiency, effectiveness, and equity through supporting more inclusive recruitment, optimizing, and offloading tasks to AI counterparts to streamline workflows, and support pay equity. While promising, ill-considered applications of AI within HR can perpetuate biases, harm worker morale, and entrench inefficiencies.

Based on the above considerations, the subcommittee recommends that UC engage in the following next steps:

1. Ensure employee representation and engagement at all phases of AI adoption, from initial analysis of whether automation should be explored and for what functions, through adoption, deployment, review, adjustment, and remedy.

2. Create a formal review process for each potential new application of AI used for HR and other purposes.

3. Create a website or other mechanism for engaging in a two-way dialogue with faculty, staff, and other affected stakeholders about the potential and actual adoption of AI in HR and its effects.

4. Conduct additional research to inform the development of a comprehensive and proactive roadmap for responsibly deploying AI in HR processes.

Each of these recommendations is explored in greater detail below.

Employee Representation & Engagement

Best practices in the integration of automation include active inclusion, fairness, a right to understand what the AI does, and access to a remedy when harms occur. Any representation in decision-making processes should include ladder-rank faculty, non-ladder-rank faculty, and other academics, as well as represented and non-represented staff, and student staff. Representatives with diverse perspectives should also be included, such as those who can speak to special needs, concerns and other EDI considerations. Such representation is essential to comply with the principles of human values, transparency, fairness/non-discrimination, shared benefit and prosperity, accuracy, and accountability—the latter especially when feedback is sought by the University on challenges with implementation, as well as seeking suggestions for appropriate accountability mechanisms and redress of harms.

Formal Review of Each Potential New Application

We recommend that campus-level councils be established to review the potential use and implementation of AI-based tools, and should assist in the development and implementation of an AI risk and impact assessment process. The campus-level councils should be tasked with determining whether a threshold should be established for when an AI-enabled tool for HR needs to submit to a formal review process (e.g., a review process may be limited to those tools that have the potential to affect an individual's rights), which could be akin to the Institutional Review Board process for research.

We recommend that the campus-level councils work with UC-wide personnel to better ensure coordination in AI HR strategies, including the following stakeholders: Vice President for Systemwide HR; Vice Provost Academic Personnel and Programs; Academic Senate; Council of UC Staff Assemblies; Vice President for Graduate and Undergraduate Affairs; Vice Provost for Equity, Diversity and Inclusion; Chief Diversity Officer; and UC campus Disabilities Services Offices. Employee representatives from potentially affected groups should also be engaged, including from UC campus Employee Resource Groups, or those identified in consultation with the Vice Provost for Equity, Diversity and Inclusion. The councils should also consult technical experts in AI/RPA, someone from UC Legal, an IT specialist, UC systemwide and UC campus location HR subject matter experts, and others with specialized expertise that may be helpful to evaluating the benefits and risks of AI/RPA in HR.

Possible tasks for the campus-level councils include assessing potential procurement and deployment of new AI-enabled HR tools (regardless of whether the AI is central or tangential to the system). Any product being considered should undergo scrutiny for compliance with the UC Responsible AI Principles and HR legal requirements. This may require revealing some nature of the underlying algorithms and their functions, the underlying data used to train the algorithms, and what steps the developer took to minimize risks of discrimination and bias.

However, scrutiny should not be limited to the procurement process. Automation and its appropriateness require ongoing review. AI-enabled tools should be frequently assessed and repeatedly adjusted by humans for compliance with each of the UC Responsible AI Principles. This includes identifying potential effects on faculty/staff and other UC stakeholders (ranging from impact on workflow to replacement of employees—both of which will require interaction and resolution with UC's union partners) and identifying whether the technology can be used with positions covered by collective bargaining. This also requires assessing how to integrate the use and assessment of the technology with existing UC policies and procedures, and whether any updates to such policies and procedures are needed. We also recommend campuses consider adding a privacy officer and/or security officer to the review process since both issues are so acute in an AI and HR context. The principles of appropriateness, fairness/non-discrimination, privacy, and security are especially implicated by this recommendation.

Create a central repository or mechanism for engagement with faculty/staff/other affected stakeholders

We recommend that campuses consider creating a website with an online form that stakeholders can use to provide feedback to the campus-level councils. The campuses can use the website to publish information about prospective or deployed AI-enabled tools in HR (and other high-risk AI application areas, such as in health, policing, and student experience outlined in this report). Some

information could be firewalled for internal use, while other information could be public-facing. Principles implicated by this recommendation especially include transparency, human values, and accountability.

Future Research

In order to introduce and integrate the principles into campus- and UC-level governing policies, UC HR personnel should conduct a deep dive into HR processes across all stages of the employee lifecycle and run each current or anticipated use of AI against the UC Responsible AI Principles. Once the principles are finalized and adopted, the campus-level councils should review UC's Staff and Academic Personnel policies, and collective bargaining agreements, as well as procurement policies and contracts. Procurement contracts should be updated not only to address privacy and data security requirements (procurement contract Appendix DS - Data Security) but also to ensure vendors are required to review and adjust their AI on a periodic basis to mitigate bias and harm.[72]

[72] "Appendix DS - Data Security", *University of California*, Aug. 12, 2019, https://www.ucop.edu/procurement-services/policies-forms/legal-forms-current/appendix-ds-8-12-2019.pdf

Introduction

UC has been continually evaluating policing on campus and its impact upon students, staff, and faculty. These include the most current UC Campus Safety Initiative, the Presidential Task Force on Universitywide Policing, and the Faculty Senate Recommendations on UC Policing.[73] Each of these initiatives has provided a series of recommendations; however, none have specifically examined the use or potential use of AI, algorithmic decision-making, or networked technologies in campus policing.

The Subcommittee on AI and Policing was established to broadly investigate the uses of AI, algorithmic decision-making, and networked technologies across UC and in discrete use cases on campuses. The subcommittee explores use of facial recognition systems, automated license plate readers, and use of social media data in campus policing. The use cases each contain recommendations on how to appropriately implement the UC Responsible AI Principles. This section concludes with overall recommendations that should guide campuses' considerations and use of AI in policing.

Use Cases for AI in Policing

Although the current use cases within UC are limited, many large-scale databases and AI-driven technologies are already used in local and national policing and may be considered by campus police departments, including University of California Police Departments (UCPD). Many of these tools were adopted without consideration of their objectives or effects and can exacerbate or amplify existing racial disparities in policing that have rightly become the focus of a national conversation. Given the civil liberties implicated and the examples of policing abuses nationwide, the use of these tools in policing at UC should be especially scrutinized. The UC Presidential Working Group on AI recommends in this report the development of a permanent council to review best practices related to the use of AI within UC. The council should actively assist UC in evaluating (and possibly retiring) AI-enabled technology used or being considered for campus policing.

In the following use cases, we seek to demonstrate the breadth and depth of the concerns that arise at the intersection of AI and policing. While recognizing the current complex and at times strained

[73] "UC Campus Safety Symposium," University of California Office of the President, March 24, 2021, https://www.ucop.edu/community-safety-plan/safety-symposium-part-two.html; "Presidential Task Force on Universitywide Policing," University of California Office of the President, 2020, https://www.ucop.edu/policing-task-force/; University of California Academic Senate, "Recommendations for UC Policing," June 29, 2020, https://senate.universityofcalifornia.edu/_files/reports/kkb-jn-recommendations-uc-policing.pdf

relationship between policing and citizens, particularly communities of color, the subcommittee considered how AI technologies are currently or may be incorporated into various aspects of campus policing. We identified three application areas: (1) Deter, (2) Prevent, and (3) Investigate.

Deter: AI-enabled tools can be used to deter crime through, for example, visible surveillance cameras with automated technology to detect criminal activity. Additionally, police body- and vehicle-cameras can reduce police misconduct and false claims of police misconduct.

Prevent: AI-enabled tools can be used to guide police patrols to high-risk areas, provide early warnings of criminal activity based on video surveillance, and, through automated facial recognition, identify individuals on campus who have previously been issued no-trespass orders.

Investigate: The use of camera surveillance and automated facial recognition can be used to analyze footage of crimes in order to identify or exonerate suspects. Additionally, automated, large-scale analysis of social media posts can uncover evidence of past crimes.

As indicated above, AI is transforming policing in many ways. Risks of its misuse require thoughtful considerations of appropriate governance approaches. In this section, we provide a more detailed look at how the UC Responsible AI Principles can be operationalized in three use cases: (1) facial recognition, (2) automated license plate readers, and (3) social media.

Facial Recognition

Facial recognition software is used to identify a person's identity from an image or video based on their facial features. Most computational algorithms work in three basic steps: (1) face detection: the location of one (or more) faces is identified in an image or video frame; (2) face analysis: a compact, numerical representation is extracted from the detected face; and (3) face matching: the extracted numerical representation is compared against a database of faces for similarity.

There currently is a moratorium in California on the use of facial recognition software by police departments.[74] Although no UC locations are currently using facial recognition software, at least two locations had considered its use before the moratorium. Outside of California, however, many local and campus police departments continue to use the nascent technology despite the various concerns regarding the technology's efficacy or the effects on civil liberties and civil rights. Campuses that contract services with local law enforcement, like UCLA and the Los Angeles Police Department, should be studied further to determine the degree to which some uses of facial recognition are being outsourced to third party vendors.

Proponents of automated facial recognition contend that this technology can be helpful in fighting crime and protecting victims. Facial recognition software developed by Thorn, for example, is currently in use by officers in all 50 US states and Canada.[75] This software has helped to identify over 17,000 child victims of human trafficking. Similarly, airport security and border control use facial recognition to identify alleged terrorist suspects.

[74] California State Assembly, "AB-1215 Law enforcement: facial recognition and other biometric surveillance." Accessed Aug. 1, 2021. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB1215

[75] "Spotlight helps finds kids faster," Thorn, accessed Aug. 1, 2021, https://www.thorn.org/spotlight/

Despite these successes, the accuracy of facial recognition, particularly for racial and ethnic minority groups as well as LGBTQ+ people, has long been of concern.[76] A 2019 study by researchers at the National Institute for Standards and Technology (NIST) reported "empirical evidence for the existence of demographic differentials in the majority of the face recognition algorithms we studied."[77] The researchers found higher false-positive rates for Asian and African American faces compared to Caucasian faces for one-to-one matching in which a photo is matched against a different photo of the same person in a database. This study verified results found by Buolamwini (2017) that Black women were the most likely to be mis-identified by commercial facial recognition, the type sold to law enforcement agencies.[78] Depending on the face-recognition algorithm, the false positive rates for one-to-one matches were 10 to 100 times higher for Asians and African-Americans than for Caucasians. Rep. Bennie G. Thompson, (D-Miss.), the chairman of the Committee on Homeland Security, said the report shows the technology is "more unreliable and racially biased than we feared."[79] In 2018, The American Civil Liberties Union (ACLU) also found evidence of racial bias in facial recognition technology when applied to African American faces.[80]

Citing these types of biases in facial recognition technology, in October 2019, California's Governor Newsom signed into law a three-year moratorium on the use of face recognition in police body cameras (this restriction took effect in January 2020). AB 1215 specifically prohibits "a law enforcement agency or law enforcement officer from installing, activating, or using any biometric surveillance system."[81]

In early 2019, the UCLA campus took up debate on UCLA Policy 133: Security Camera Systems.[82] This policy aimed to centralize and standardize the collection of data from the some 3,000 cameras on the UCLA campus. A critical component of comprehensive safety, it is argued in this policy, is the use of security camera systems designed to deter crime, aid in the apprehension of suspects, and enhance the overall safety and security of property and individuals of the campus community. Although facial recognition was not specifically part of the policy, the issue of automated facial recognition was discussed. Proponents argued that facial recognition could be an effective tool to determine if those with a stay-away order were on campus. Opponents, however, raised concerns of privacy and data storage, management, and security.

In February 2020, UCLA announced it would abandon plans to use facial recognition technology on its campus. Vice Chancellor Michael Beck said the university "determined that the potential

---

[76] Larry Hardesty, "Study finds gender and skin-type bias in commercial artificial-intelligence systems," *MIT News*, February 11, 2018, https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212.

[77] "NIST study evaluates effects of race, age, sex on face recognition software," *NIST*, Dec. 19, 2019, https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software.

[78] Joy Buolamwini, "Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers" (Master's Thesis, MIT, 2017).

[79] "Government facial recognition report confirms unreliability and racial bias," Statement from Rep. Bennie G. Thompson (D-MS), Chairman on Homeland Security, https://homeland.house.gov/news/press-releases/government-facial-recognition-report-confirms-unreliability-and-racial-bias-.

[80] Jacob Snow, "Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots." *ACLU*, July 26, 2018, https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28 .

[81] ACLU Southern California, "The Body Camera Accountability Act (AB1215)," 2019, https://www.aclusocal.org/en/legislation/body-camera-accountability.

[82] "UCLA Policy 133: Security camera systems DRAFT for public review," N.D., http://www.adminpolicies.ucla.edu/pdf/133-DRAFT-2018-09-05.pdf.

benefits are limited and are vastly outweighed by the concerns of the campus community."[83] These risks were highlighted by opponents who used Amazon's facial recognition software, Rekognition, to compare publicly available photos of UCLA athletes and faculty against mugshots. Of the 400 photos analyzed, 58 returned false positives with 100% confidence.[84]

At the same time, law enforcement agencies around the country continue to use facial recognition technologies without proper training or supervision. Facial recognition software provided by the firm Clearview AI, for example, is being used by thousands of police officers and government employees across the country. Between 2018 and 2020, this facial recognition software has been deployed in over 340,000 searches at 1,803 unique public agencies.[85] Troublingly, it has been reported that some investigators have used this facial recognition software without the knowledge of their supervisors.

> **"Keeping our students, faculty, and staff safe is and should be one of our highest priorities."**

In addition to these concerns, Clearview AI grew its searchable database of more than 3 billion images—without permission, and in violation of terms of service—by scraping social media sites including Facebook, Google, Instagram, LinkedIn, and Twitter, among others. Senator Chris Coons (D-Del.), the chair of the Senate Judiciary Subcommittee on Privacy, Technology, and the Law, stated: "We have little understanding of who is using this technology or when it is being deployed. Without transparency and appropriate guardrails, this technology can pose a real threat to civil rights and civil liberties."[86]

Because of the ubiquity of surveillance cameras, police body-cams, and citizen-generated imagery, issues of privacy, accuracy, and bias in facial recognition should be carefully considered within UC. To this end, this subcommittee heard from a representative from the UCSD Campus Police and learned that the UCSD campus has experimented with facial recognition, in addition to other video-based surveillance technologies.[87] A centralized video surveillance management system is used on the UCSD campus for the purposes of detection, mitigation, and historical archival of criminal activity. The subcommittee learned that UCSD experimented with facial recognition as well as other forms of hard (i.e., highly distinct characteristics like faces) and soft (e.g., less distinct characteristics like gait) biometric identification.

Surprisingly, no formal protocols appear to be in place at UCSD to determine the suitability or appropriateness of deploying new technologies, nor any formal mechanism for review of existing technologies. Notwithstanding the current moratorium, we recommend that all campuses institute a formal policy for assessing whether and if to deploy AI-enabled policing technologies and ensure, if deployed, they are subject to continuous oversight and evaluation.

---

[83] Edward Ongweso Jr., "UCLA abandons plans to use facial recognition after backlash," *Vice*, Feb. 19, 2020, https://www.vice.com/en/article/z3by79/ucla-abandons-plans-to-use-facial-recognition-after-backlash.

[84] Ibid.

[85] Ryan Mac, Caroline Haskins, Brianna Sacks, & Logan McDonald, "Surveillance Nation," *Buzzfeed News*, April 9, 2021, https://www.buzzfeednews.com/article/ryanmac/clearview-ai-local-police-facial-recognition.

[86] Ibid.

[87] Interview conducted by Policing Subcommittee member of a member of the UCSD campus community, 2021.

Keeping our students, faculty, and staff safe is and should be one of our highest priorities. UC should not deploy AI-enabled tools without fully considering potential benefits and risks. This includes taking into account accuracy and bias to inform appropriate development and deployment strategies and implementation of strict data controls to minimize risks to privacy and civil liberties. Attention should be paid to how data is collected, stored, used, and shared with third parties (particularly those seeking data to train their next generation of AI-enabled tools).

Principles Implicated & Sub-Recommendations:

- **Appropriateness**: Before deploying any facial recognition technology, UC should evaluate its benefits and risks in comparison to other options, specify the scope of appropriate uses and prohibit uses that are incompatible with UC's stated values including suspicionless monitoring of persons on campus. Because of the particularly sensitive nature of recording and storing biometric facial information, and in alignment with AB 1215, we recommend a three-year moratorium on the use of facial recognition in all policing situations. Any future deployment of facial recognition should be carefully balanced against civil liberty and civil rights concerns.

- **Transparency**: If facial recognition is deployed, its deployment and use must be made transparent to all members of the UC community, including how it is being used, how any recorded data is stored, and if any other parties will be granted access to recorded data. Analogous to the proposed federal repository, a University-wide repository enumerating all deployed technologies should be created, allowing for an easier assessment and mitigation of risks associated with facial recognition.[88] Because of the ease with which some facial recognition technologies can be downloaded and used on individual devices, we recommend a ban on the use of these technologies by individual officers or individual departments.

- **Accuracy/Fairness/Bias**: Given the significant fairness and bias concerns raised by current facial recognition software, any future facial recognition software must be carefully and thoroughly evaluated for accuracy, fairness, and bias. These evaluations should not be left to software vendors but rather conducted by independent evaluators in real-world scenarios.

- **Privacy**: While facial recognition can be a useful tool for ensuring safety on our campuses, its benefits must be weighed against individual privacy and civil liberties. Because of the complexity of balancing these somewhat competing interests, we recommend that if facial recognition is deployed, it be done in incremental, manageable steps, allowing for frequent assessment and reevaluation of the relative advantages and disadvantages.

- **Human Values:** Given the ways racism has shaped policing, the use of AI-enabled tools in policing must be reviewed to avoid furthering bias, undermining human dignity, interfering with the exercise of civil rights, and violating human rights. Alternatives to AI-enabled policing, including decommissioning AI-enabled technologies, should be considered.

[88] "Facial recognition technology: Federal law enforcement agencies should better assess privacy and other risks," *GAO*, June 2021, https://www.gao.gov/assets/gao-21-518.pdf

- **Shared Benefit and Prosperity:** AI-enabled policing should be aimed at bringing equitable benefits to all. This requires attention to who is subject to and who benefits from any proposed use of AI-enabled policing tools.

Automated License Plate Readers

Automated license plate readers (ALPR) have been used in American police departments since the 1990s. These systems typically involve either fixed or mobile cameras that capture video or still images of passing vehicles.[89] An algorithm detects the numbers and letters in the captured image and "reads" the license plate. Most ALPR systems then store the information about the captured image, including not only the license plate, but also the time and GPS coordinates of each scan. Extensive collection of ALPR data on individuals permits the tracking of their movements in space and time.[90] A third step typically involves the comparison of captured images with "alert" or "hot" lists (i.e., a database of sought-after vehicles). These lists can include license plates associated with stolen vehicles, vehicles suspected of being used in criminal activity, or owned or associated with wanted or missing persons.[91] A positive match might prompt police reaction.

Like many forms of policing technologies, ALPR use can, in theory, represent the use of AI with the promise of significant benefits. As the software for license plate recognition improves and as equipment and data storage costs plummet, ALPR can provide a "force-multiplier" effect for police departments.[92] Like other forms of video analytics, ALPR might cheaply and efficiently increase the capabilities of the police. Plate readers can identify vehicles associated with past or ongoing criminal activity.[93] And some high-profile individual uses of ALPR have met success, such as the 2014 apprehension of a person suspected of multiple highway shootings in Kansas City, Missouri.[94]

At the same time, the potential concerns raised by ALPR are varied and significant. Today ALPR use can be characterized as widespread, underregulated, and raising significant civil liberties concerns. Each of these considerations bear on the use of ALPR as an AI application within UC.

First, ALPR technology is ubiquitous. Law enforcement agencies around the country use tens of thousands of readers. A 2015 Justice Department report noted that more than three quarters of

---

[89] "Second report of the Axon AI & Policing Technology Ethics Board: Automated license plate readers," *Axon AI*, Oct. 2019, https://static1.squarespace.com/static/58a33e881b631bc60d4f8b31/t/5dadec937f5c1a2b9d698ba9/1571679380452/Axon_Ethics_Report_2_v2.pdf

[90] Julia Angwin and Jennifer Valentin-Devries, "New tracking frontier: Your license plates," *The Wall Street Journal*, Sept. 29, 2012, https://web.archive.org/web/20130124093854/http://online.wsj.com/article/SB10000872396390443995604578004723603576296.html

[91] David J. Roberts and Meghann Casanova, "Automated license plate recognition systems: Policy and operational guidance for law enforcement," *Department of Justice, National Institute of Justice*, 2012, https://www.theiacp.org/sites/default/files/IACP_ALPR_Policy_Operational_Guidance.pdf

[92] Tim Simonite, "AI license plate readers are cheaper—so drive carefully," *Wired*, Jan. 27, 2020, https://www.wired.com/story/ai-license-plate-readers-cheaper-drive-carefully/; Keith Gierlack, Shara Williams, Tom LoTourrette, James M. Anderson, Lauren A. Mayer, & Johanna Zmud, "License plate readers for law enforcement: Opportunities and obstacles," *Rand Safety and Justice Program*, 2014, https://www.ojp.gov/pdffiles1/nij/grants/247283.pdf

[93] John S. Hollywood, Michael J.D. Vermeer, Dulani Woods, Sean E. Goodison, & Brian A. Jackson, "Using video analytics and sensor fusion in law enforcement," *Priority Criminal Justice Initiative, A project of the RAND Corporation, the Police Executive Research Forum, RTI International, and the University of Denver*, 2018, https://www.rand.org/content/dam/rand/pubs/research_reports/RR2600/RR2619/RAND_RR2619.pdf

[94] Matt Pearce, "How technology helped crack the Kansas City highway shooter case," *LA Times*, April 20, 2014, https://www.latimes.com/nation/nationnow/la-na-nn-kansas-city-highway-shooter-20140419-story.html

police departments serving populations of 100,000 or more residents used ALPR technology.[95] Some of these cameras can scan nearly 2000 plates per minute.[96] Several UC campuses currently use ALPR readers and store license plate data, including UC Irvine, UCLA, UC Merced, UC Riverside, UC San Diego, and UC Santa Barbara.[97]

In addition to directly collecting license plate data, police departments can also indirectly access it from private sources. Private companies like Vigilant that store ALPR data collected by repossession agents can also sell that information to willing law enforcement agencies.[98] In addition, the availability of private license plate reader networks like Flock Safety allow neighborhoods to collect ALPR data, which can be voluntarily shared with police departments.[99] As a result, police departments can access billions of plate scans already stored in databases with more added every day.

Second, the use of ALPR technology and retention of the resulting data is widely considered to be underregulated. No federal law explicitly regulates ALPR technology, and a 2021 survey found only 16 states with any laws on ALPR use.[100] In California, there are civil codes that pertain to ALPR technology and related data: Calif. Vehicle Code § 2413 puts in place requirements for ALPR data retention, use, and reporting for California Highway Patrol and Calif. Civil Codes § 1798.29 and 1798.90.5 where APLR data is defined as personally identifiable information (PII) for breach notification purposes.[101] With respect to the Fourth Amendment's prohibition on unreasonable searches and seizures, some have argued that recent Supreme Court decisions like Carpenter v.

---

[95] Brian A. Reaves, "Local police departments, 2013: Equipment and technology," *US Department of Justice*, July 2015, https://bjs.ojp.gov/content/pub/pdf/lpd13et.pdf

[96] Angel Diaz and Rachel Levinson-Waldman, "Automatic license plate readers: Legal status and policy recommendations for law enforcement," *Brennan Center for Justice*, Sept. 10, 2020, https://www.brennancenter.org/our-work/research-reports/automatic-license-plate-readers-legal-status-and-policy-recommendations.

[97] University of California Irvine Police Department, "Policy 403: Automated license plate readers," https://www.police.uci.edu/how-do-i/_img/automated_license_plate_readers; University of California Los Angeles, "UCLA Policy 134: Automated license plate recognition system and information DRAFT for public review," http://www.adminpolicies.ucla.edu/pdf/134-DRAFT-2019-04-17.pdf; University of California Merced, "Automated license plate recognition procedure," https://taps.ucmerced.edu/sites/taps.ucmerced.edu/files/documents/lprprocedures_2.pdf; University of California Riverside Transportation Services, "License plate recognition," https://transportation.ucr.edu/lpr; "The University of California San Diego awarded Park Assist the parking guidance system for its new Osler Parking Structure," *Park News*, Jan. 7, 2019, https://blog.parknews.biz/2019/01/the-university-of-california-san-diego-awarded-park-assist-the-parking-guidance-system-for-its-new-osler-parking-structure/; University of California Santa Barbara Transportation & Parking Services, "UCSB ePermits launching July 1, 2021," 2021, https://www.tps.ucsb.edu/news/introducing-automated-license-plate-recognition-alpr.

[98] Joseph Cox, "This company built a private surveillance network. We tracked someone with it." *Vice*, Sept. 17, 2019, https://www.vice.com/en/article/ne879z/i-tracked-someone-with-license-plate-readers-drn; Tim Cushing, "Private companies gathering plate data are selling access to people's movements for $20 a search," *TechDirt*, Sept. 23, 2019, https://www.techdirt.com/articles/20190918/14224343020/private-companies-gathering-plate-data-are-selling-access-to-peoples-movements-20-search.shtml; Joseph Cox, "Customs and border protection bought access to nationwide car tracking system," *Vice*, July 17, 2020, https://www.vice.com/en/article/qj4zgm/customs-border-protection-cbp-license-plate-reader-database.

[99] Alfred Ng, "License plate tracking for police set to go nationwide," *MSN News*, Aug. 18, 2020, https://www.msn.com/en-us/news/technology/license-plate-tracking-for-police-set-to-go-nationwide/ar-BB187gbb

[100] "ALPR FAQs," *The IACP*, Aug. 18, 2018, https://www.theiacp.org/resources/alpr-faqs; "Automated license plate readers: State statutes," *National Conference on State Legislature*, April 9, 2021, https://www.ncsl.org/research/telecommunications-and-information-technology/state-statutes-regulating-the-use-of-automated-license-plate-readers-alpr-or-alpr-data.aspx.

[101] Calif. Vehicle Code § 2413 (2011), https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=VEH&sectionNum=2413; Calif. Civil Code § 1798.29 (2020), https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.29; Calif. Civil Code § 1798.90.51 (2016), https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=1798.90.51.&lawCode=CIV.

United States (2018) may provide some restrictions.[102] However, the Court has not yet decided a case directly addressing the constitutionality of ALPR. Under current Fourth Amendment law, the government's justification of capturing license plate data without warrants or other judicial safeguards rests on the presumption that people lack any reasonable expectation of privacy in that information.[103] Just as important, any Fourth Amendment restriction would apply only to direct collection of ALPR data by police and not at all to the government's purchase of that data.

The paucity of regulation means that many of the policy questions surrounding ALPR use remain unresolved, despite the growing size of ALPR data accessible to police departments, including UC police. One fundamental question underlying the value of ALPR is its ability to effectively address crime. What "hit rates" do police departments experience? Available research provides at best a mixed picture, suggesting that even with very large license plate databases, successful identification of wanted vehicles was low.[104] And when the reasons for creating "hit lists" are unregulated, government agencies might use them for anything, including trying to flag uninsured drivers to fine them.[105]

Additionally, when there is little regulation about retaining ALPR data and for what purposes, these decisions are left to law enforcement agencies or other government entities to determine. A high degree of police discretion raises numerous concerns. Police may engage in improper or abusive practices in tracking persons. ALPR technology may be used heavily in some communities and not others, resulting in surveillance with racially disparate effects.[106] Police may also choose to deploy ALPR against those who are lawfully exercising their First Amendment rights in peaceful protests.[107] These concerns are all magnified by the problem of inaccuracies and mistakes in ALPR technology. A misread license plate, coupled with highly discretionary decisions by police officers, can lead to innocent drivers being pulled over and having guns drawn at them.[108]

The current atmosphere of ALPR use by US police departments might be characterized as a practice that lacks substantial regulation, engages in mass data collection, and raises significant privacy and other civil liberties concerns.[109] In particular, the lack of attention to basic matters such as data retention, data access, permissible uses, and accountability and transparency measures remains important and concerning.

---

[102] Angel Diaz and Rachel Levinson-Waldman, "Automatic license plate readers: Legal status and policy recommendations for law enforcement," *Brennan Center for Justice,* Sept. 10, 2020, https://www.brennancenter.org/our-work/research-reports/automatic-license-plate-readers-legal-status-and-policy-recommendation.

[103] Indeed, ALPR policies in place at UCLA and UC Merced state that ALPR use is allowed for this reason. See "UCLA Policy 134: Automated License Plate Recognition Systems and Information," *UCLA Events and Transportation*, 2019, http://www.adminpolicies.ucla.edu/pdf/134-DRAFT-2019-04-17.pdf; "Automated License Plate Recognition Procedure," *UC Merced Transportation and Parking Services*, n.d., https://taps.ucmerced.edu/sites/taps.ucmerced.edu/files/documents/lprprocedures_2.pdf.

[104] George Joseph, "What are license-plate readers good for?" *Bloomberg,* Aug. 5, 2016, https://www.bloomberg.com/news/articles/2016-08-05/license-plate-readers-catch-few-terrorists-but-lots-of-poor-people-of-color

[105] "On camera: New statewide program to catch uninsured drivers," *The Shawnee News-Star*, Oct. 24, 2018, https://www.news-star.com/news/20181024/on-camera-new-statewide-program-to-catch-uninsured-drivers

[106] Dave Mass and Jeremy Gillula, "What you can learn from Oakland's raw ALPR data," *Electronic Frontier Foundation,* Jan. 21, 2015, https://www.eff.org/deeplinks/2015/01/what-we-learned-oakland-raw-alpr-data

[107] Caroline Haskins and Ryan Mac, "Here are the Minneapolis Police's tools to identify protestors," *Buzzfeed News*, May 29, 2020, https://www.buzzfeednews.com/article/carolinehaskins1/george-floyd-protests-surveillance-technology

[108] Charlie Warzel, "When license-plate surveillance goes horribly wrong," *The New York Times*, April 23, 2019, https://www.nytimes.com/2019/04/23/opinion/when-license-plate-surveillance-goes-horribly-wrong.html.

[109] Lauren Feiner and Annie Palmer, "Rules around facial recognition and policing remain blurry," *CNBC,* June 12, 2021, https://www.cnbc.com/2021/06/12/a-year-later-tech-companies-calls-to-regulate-facial-recognition-met-with-little-progress.html.

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** Many UC campuses already use ALPR technology for seemingly mundane matters like parking, but ALPR can also be used in a variety of ways that might implicate civil liberties, free speech, and privacy. UC should specify the scope of appropriate uses for ALPR and prohibit uses that are incompatible with University values, including suspicionless monitoring of persons engaged in First Amendment–protected activities.

- **Transparency:** UC must provide accessible information about the deployment of ALPR, the kinds of data collected by ALPR and similar AI-enabled tools, the uses for which that data is collected, and those actors and institutions with access to the data. This transparency should take the form of regular and public reporting to the maximum extent possible.

- **Privacy:** Any uses of ALPR and similar AI technologies should be guided by a respect for privacy, including but not limited to locational and associational privacy. In accordance with Calif. Civil Codes § 1798.29 and 1798.90.51, UC must maintain written policies on ALPR data collection, use, and oversight processes (i.e., data privacy practices in compliance with relevant laws) that are publicly accessible and subject to regular review.[110] UC community members should be granted the right to know what information exists about themselves in policing databases, upon request.

- **Accountability:** UC should establish accountability mechanisms that would respond to inappropriate or abusive uses of data collected by ALPR or similar technologies that may be intentional, reckless, or negligent. Such accountability mechanisms should consider both individual remedies as well as systemic responses when such misuses occur.

- **Shared Benefit and Prosperity:** Any use of ALPR should be reviewed for equity. This should include a review of its performance, as well as why, where, and when it is deployed.

Social Media

Social media networks are "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded digital system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system."[111]

Social media have been used to investigate crimes, including those occurring on campus. An early example of the use of social media for campus police work comes from the University of Colorado in 2006.[112] In an effort to stop an annual student gathering for mass marijuana consumption on April 20, campus police fenced the field students intended to use and lined its perimeter with surveillance cameras, posted "no trespassing" signs, and warned students that the area was under surveillance. Undeterred, students clambered over the fence and, as they had in many prior years, smoked

[110] Calif. Civil Code § 1798.29 (2020), https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.29; Calif. Civil Code § 1798.90.51 (2016), https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=1798.90.51.&lawCode=CIV

[111] D. Boyd and N.B. Ellison, 2007. "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer Mediated Communication*, 13 (1), 210–230. 10.1111/j.1083-6101.2007.00393.

[112] Ryan Shaw, "Recognition markets and visual privacy." *UnBlinking: New Perspectives on Visual Privacy in the 21st Century*, https://www.law.berkeley.edu/files/bclt_unblinking_shaw.pdf.

marijuana. The campus police took stills from the video footage of the event, posted images of students smoking on their web site, and offered fifty-dollar rewards to individuals who provided identifying information. By crowdsourcing recognition of individuals, the police were able to identify a substantial number of students and charge them with trespassing. In doing so, the campus police raised serious privacy and security concerns.

Today, campus police are increasingly turning to social media platforms to identify individuals engaged in a range of activities, from First Amendment-protected protests or illegal behavior.[113] But, more significantly, a number of companies now aggregate and analyze data from social media and other sources and sell tools to law enforcement agencies. These tools are marketed to universities, colleges, and K-12 schools as a means to identify social media postings that may indicate suicidal ideation or presage violence. They have also been used to monitor students during lawful protests.[114]

Companies such as Palantir, Social Sentinel, and Banjo ingest social media profiles, posts, and metadata (e.g., text, images, location data, and network connections) and use them to identify patterns, make predictions, and surface information for action. AI-enabled social media surveillance may use network analysis to identify relationships between individuals, natural language processing and sentiment analysis to classify and assign meaning to posts, and use location traces to identify individuals' movements and associations and predict future ones. Some social media surveillance products combine online data with other information garnered from public and private sources.[115] Research finds that "the shift from traditional to big data surveillance is associated with a migration of law enforcement operations toward intelligence activities."[116] Of particular concern with respect to social media surveillance and ALPR data (described in the previous use case), the use of AI-enabled big data techniques is bringing personal data and communications from disparate spheres of individuals' lives, many of whom have no prior contact with the criminal justice system, into the view and at times into the databases of law enforcement.[117]

There is evidence that UC campus police are making use of these types of tools. For example, police responding to the student strikes at UC Santa Cruz in 2020 apparently had access to the LEEP (Law Enforcement Enterprise Portal) data-sharing platform.[118] Anecdotal reports from journalists include students recounting how police officers seemed to be able to identify them and knew personal information about them. However, UC Santa Cruz stated the trackers were used only "to know the location of on-duty officers who were helping keep people safe' and "there was no tracking of students or strikers."[119] Companies like Dataminr have been key partners to law enforcement and found to be actively racially-profiling the public, including participants in Black

---

[113] For example see, Thorburn, Elise Danielle. "Social media, subjectivity, and surveillance: Moving on from occupy, the rise of live streaming video." *Communication and Critical/Cultural Studies* 11.1 (2014): 52-63.
(describing the use of social media live-streaming footage by the police who "can watch the stream live...observe and record images, determine who is a frequent participant based on whose images are seen regularly, direct officers to scenes that the live stream crew may be filming, and select individuals for the officers to target, all from a secure and distant location).

[114] Ari Sen, "UNC campus police used geofencing tech to monitor antiracism protestors," *NBC news*, December 21, 2019, https://www.nbcnews.com/news/education/unc-campus-police-used-geofencing-tech-monitor-antiracism-protestors-n1105746

[115] For an example of a contract which included work for the University of Utah see, State of Utah contract with Banjo" (PDF). utah-das-contract-searchsp.s3.amazonaws.com.

[116] Sarah Brayne. "Big data surveillance: The case of policing." *American sociological review* 82.5 (2017): 977-1008.

[117] Ibid.

[118] Nicole Karlis, "Emails show UC Santa Cruz police used military surveillance to suppress grad student strike," *Salon*, May 18, 2020, https://www.salon.com/2020/05/18/emails-show-uc-santa-cruz-police-used-military-surveillance-to-suppress-grad-student-strike/

[119] Ibid.

Lives Matter protests in 2020.[120] UC-supported predictive policing software, PredPol, has also come under increased scrutinty within the UC community for its use in racial profiling by law enforcement.[121]
Social media surveillance by police continues to increase. A 2014 report found that police reported increased use of social media for investigations, with Facebook, YouTube, and Twitter being the most common.[122] Common uses of social media surveillance for investigations include identifying and locating suspects through social media posts, photos, and "friends of friends." Police also report using social media surveillance for planning and response including preparing for First Amendment-protected public gatherings and protests, preventing or thwarting crime, executing warrants, and tracking alleged gang behavior. They also report using social media to enlist users' assistance, for example soliciting information from the public about ongoing or unsolved crimes. In recent years, sentiment analysis and image recognition tools are allowing police to make use of social media on a larger scale. Today, AI-driven social media surveillance tools and services aggregate, organize, and analyze personal information and communications collected from multiple social media platforms.

Some forms of social media surveillance—for example, an individual officer looking for a social media profile—require no university investment in infrastructure or services. Police may access information through their own accounts or through search engines to the extent it is available. This reduces visibility into some forms of social media surveillance by police and may make them particularly difficult to control. The social media surveillance tools and services that provide predictive analytics or other population-level analysis discussed above are likely to require a campus- or UC-level investment, making them more visible and controllable through policy and oversight. As in general policing, the use of social media surveillance on campus is poorly documented and underregulated.

The use of social media surveillance, along with facial recognition, ALPR, and other big-data policing techniques fundamentally change police surveillance because they rely on "the inter-institutional integration of data and proliferation of dragnet surveillance practices—including the use of data on individuals with no direct police contact and data gathered from institutions typically not associated with crime control."[123]

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** The suspicionless monitoring and the review and/or collection of data on individuals who have no prior interaction with police and no reason to be under surveillance is inappropriate. The UC police should not use social media data to carry out its activities.

- **Transparency:** The lack of visibility into the data collection practices of vendors, some of whom are known to collect data through improper and potentially illegal methods (e.g., Banjo used Trojan apps to scrape content from unwitting users' social media accounts), is

[120] Sam Biddle. "Twitter Surveillance Startup Targets Communities of Color for Police," *The Intercept*, Oct. 21, 2020, https://theintercept.com/2020/10/21/dataminr-twitter-surveillance-racial-profiling/.
[121] Leila Miller. "LAPD Will End Controversial Program That Aimed to Predict Where Crimes Would Occur," *Los Angeles Times*, April 21, 2020, https://www.latimes.com/california/story/2020-04-21/lapd-ends-predictive-policing-program.
[122] *Social Media Use in Law Enforcement: Crime prevention and investigative activities continue to drive usage*, LexisNexis Risk Solutions, November 2014, https://centerforimprovinginvestigations.org/wp-content/uploads/2018/11/2014-social-media-use-in-law-enforcement-pdf.pdf.
[123] Sarah Brayne, "Big data surveillance: The case of policing," in *The Cambridge Handbook of Policing in the United States*, ed. Tamara Rice Lave and Eric J. Miller (Cambridge University Press, 2019), 511-530.

concerning. Any UC police activities using social media data must clearly communicate this use.

- **Privacy and Security:** The scope of the data that is collected and made available may be far beyond the minimum necessary to achieve particular law enforcement goals. Due to these risks, we do not recommend UC police utilize social media data in its activities.

- **Human Values:** While social media may provide useful information for investigating particular criminal acts, general social media surveillance is inconsistent with the exercise of civil liberties and civil rights.

Recommendations on Implementation of UC AI Principles in Policing

Campus police are increasingly turning to AI-enabled technologies to support greater efficiency and effectiveness in policing, including use of facial recognition, automated license plate readers, and social media data in an attempt to deter, prevent, and investigate crimes. While these applications may seem promising, they simultaneously pose serious risks to privacy, security, and civil liberties. As such, we recommend that UC engage in the following next steps:

1. Ensure appropriate oversight and assessment prior to the procurement and use of AI-enabled tools in policing, and consideration of decommissioning AI-enabled technologies used in policing that are inconsistent with the UC Responsible AI Principles.

2. Develop a database that inventories and tracks all uses and versions of AI-enabled tools used in policing, including system updates, vendors, documentation, and disclosures.

3. Perform periodic review of all forms of AI used in policing and require all vendors to provide explainability and transparency of said technologies, including publicly disclosing data used.

Each of these recommendations is explored in greater detail below.

Oversight and Assessment

Every current and prospective use of AI for policing should be evaluated for its appropriateness for meeting UC-wide objectives, legal compliance requirements, and adherence to the UC Responsible AI Principles. For example, does a specific AI-enabled tool provide an effective and equitable mechanism for ensuring safety that could not be otherwise achieved? And, if so, is any benefit from this AI-enabled tool offset by concerns of fairness, bias, and privacy? Careful consideration should be given to the potential tradeoffs that AI-enabled tools can bring for efficiency in police operations with their implications for effectiveness and equity. In order to do so, UC should establish explicit criteria for determining appropriate levels of accuracy related to potential risk(s) posed to individuals and the UC community and, at minimum, procedures for identifying and remedying bias or discrimination in datasets and models. Policies and training should prohibit police from using or adopting AI-enabled tools that have not been assessed and approved.

Development of a Database

To support transparency and accountability, a public database that clearly inventories and tracks all uses and versions of AI-enabled tools used in policing, including system updates, vendors,

documentation, and disclosures, should be established. To the extent possible, the database should also document data used, including details on how data is collected and any underlying assumptions in its collection and use. This recommendation is in alignment with the recommendations put forward by the UC Presidential Task Force on Universitywide Policing and the UC Community Safety Plan to increase transparency in UC police activities.[124] Additionally, a process should be established that allows members of the UC community to provide feedback on the AI-enabled tools and contest or decommission their use.

Periodic Review

In addition to the development of a database to document use of AI in policing, procedures must be put in place to support regular review of AI-enabled tools implemented. Particular consideration should be given to the development of effective evaluation strategies, such as implementation of risk and impact assessments, to evaluate the accuracy, fairness, and bias of AI-enabled tools over time. Based on these assessments, procedures must be established that allow for the halting or discontinuation of a tool and implementation of appropriate remedial measures for problematic uses and outcomes.

[124] "University of California Presidential Task Force on Universitywide Policing: Implementation Report," *University of California Office of the President*, June 2020, https://www.ucop.edu/policing-task-force/uptf-final-implementation-report_june-2020.pdf;"UC Community Safety Plan," *University of California Office of the President*, 2021, https://www.ucop.edu/community-safety-plan/_files/uc-community-safety-plan.pdf

Introduction

In addition to research and public service, a primary pillar of UC's mission is teaching and educating students. Numerous AI applications have emerged in recent years to assist University staff, faculty, and students to improve metrics of student success. Given the potential benefits and risks of AI, the UC Presidential Working Group on AI formed a subcommittee to provide guidance on whether and how UC can responsibly deploy such applications across multiple facets of the student experience.

The student experience subcommittee has examined the current and potential uses of AI-driven tools and platforms in connection with students at UC campuses. We use the term "student experience" to refer to the entire lifecycle of a student's interactions with the University, from the time students apply for admission and financial aid, through enrollment and their student careers. We identify four specific domains in which AI may be relevant:[125]

- Admissions and Financial Aid
- Retention, Student Advising, and Academic Progress
- Student Mental Health and Wellness
- Grading and Remote Proctoring

AI-enabled tools and platforms may be especially appealing to large public universities, which recruit, admit, and educate thousands of students, often under tight budget and staffing constraints. When resources are limited, AI can assist, supplement, or accelerate human review and decision-making, and complement efforts by advising and counseling staff often stretched thin to serve students adequately. This automation can provide efficiencies that improve performance metrics of students, faculty, and staff alike. AI can also provide a check on human biases, helping to increase

[125] The committee recognizes that several of these areas fall under the purview of the Academic Senate (systemwide and on individual campuses) to establish policy. Our recommendations are meant to offer guidance regarding opportunities and cautions for implementation of AI-enabled tools under the broad category of "student experience."

fairness and accountability in institutional decision making.[126] At the same time, AI platforms raise important concerns about fairness, transparency, discrimination, accountability, privacy, and other elements of the UC Responsible AI Principles among individuals who may feel they have little choice to opt-out.

Use Cases for AI in Student Experience

The section assesses the opportunities and hazards of automation software and AI-driven higher ed applications in four domains:

- **Admissions and Financial Aid:** AI programs are available to improve admissions yield (not only in the offer of admission but also in financial aid decisions) and reduce the summer pre-enrollment gap or "summer melt" (when a student commits to a college but does not ultimately enroll).

- **Retention, Student Advising, and Academic Progress:** AI programs can provide feedback on student engagement and performance, assist academic advisors, and ultimately improve student retention, academic success, and graduation rates.

- **Student Mental Health and Wellness:** AI can help attend to the social-emotional needs of students during this critical time of their early adulthood and flag cases for human intervention as warranted.

- **Grading and Remote Proctoring:** AI can be used to automate grading, detect academic misconduct, and proctor tests remotely.

Admissions & Financial Aid

In the most recent application cycle, UC received 250,000 applications for admission to its undergraduate programs, 16% more than the previous year and an all-time high.[127] Decision-making regarding offers of admission is controlled at each individual campus through separate but similar processes of evaluation.[128] Although only minimally used at UC currently, AI-enabled tools are deployed in different aspects of the admissions and financial aid processes at both the undergraduate and graduate levels.

According to interviews with UC admissions professionals, some campuses use algorithmic models to aid in the review of applications.[129] With each UC campus receiving tens of thousands of admission applications each cycle, tasking humans with reading each application to make admissions decisions becomes increasingly challenging. Algorithms could be employed to create a predictive score for each applicant, a process that could help to supplement evaluation by admissions staff. The use of algorithms can lead to gains in efficiency, effectiveness, and equity by

[126] Scott Jaschik, "Do college application essays favor wealthier students?" *Inside Higher Ed*, June 1, 2021, https://www.insidehighered.com/admissions/article/2021/06/01/do-college-application-essays-favor-wealthier-students.

[127] "All-time record high number of applicants apply to UC," *UCOP Press Room*, January 28, 2021, https://www.universityofcalifornia.edu/press-room/all-time-record-high-number-applicants-apply-uc-chicanolatino-students-comprising-largest.

[128] Information in this section provided by admissions staff at UC Santa Barbara, UC Riverside, and UCOP.

[129] The Student Experience Subcommittee members interviewed Han Mi Yoon-Wu, Executive Director, Systemwide Undergraduate Admissions; Lisa Przekop, Director of Admissions, UC Santa Barbara; Emily D. Engelschall, UCR Director of Admissions; Michelle Whittingham, Associate Vice Chancellor for Enrollment Management, UC Santa Cruz.

revealing and/or countering potential human bias and lack of consistency (equality) in reviews. However, the potential for adverse outcomes or unintended consequences can be ingrained in AI-enabled tools if they draw on outdated training data or data that is incomplete or unrepresentative of a broad demographic. Algorithmic bias can manifest in many ways because proxies for characteristics such as race, ethnicity, and gender can be misleading (e.g., when zip code closely matches race/ethnicity). While AI can help identify and reduce the impact of human biases, it can also make the problem worse by reinforcing and deploying biases at scale in sensitive application areas.

Statistical models have long been used to some extent at several UC campuses. These are typically a weighted average but manually adjusted. The features in this average involve data points such as academic achievement (e.g., high school grades), the personal statement (typically read and scored by a human), and socioeconomic factors. The data is combined into a formula or model, prepared in-house by the institutional research team and sometimes with advice from an outside vendor. While admissions criteria for students are public, the models or formulas themselves are not, according to the UC admissions professionals the subcommittee interviewed.

> **"Adverse outcomes or unintended consequences can be ingrained in AI-enabled tools if they draw on outdated training data or data that is incomplete or unrepresentative of a broad demographic."**

Significant data cleaning of each application (e.g., tracking down misreported or blank information) is often required, particularly for fringe cases. This process can be quite time-consuming, and while students may submit the same application to multiple campuses, admissions staff at the various campuses have no centralized authority to perform this data-cleaning function for all, resulting in duplication of effort (and wasted time and human resources) across the selected campuses. Use of AI-enabled tools could streamline this process, making it more manageable for staff. Admissions offices are charged not only with selecting which candidates will be offered admission but also managing the number of acceptances (i.e., those students who return a Statement of Intent to Register (SIR)). AI models may also be used to predict this "yield," how many and under what conditions students who have been offered admission will actually enroll.[130]

Another example is financial aid. Financial aid offices are involved not only in crafting initial aid packages offered upon admission but also often need to make instant decisions on aid to some of the most at-risk students to ensure they have sufficient funding to finish a semester. A quick decision based on an algorithm can help students get timely help and allow staff to focus more on informing borrowing decisions and helping students understand the amount being borrowed, options to reduce student debt, and ensuring sustainability of their funding. AI could help increase degree completion by ensuring that students are aware of all of their funding options and the obligations they will incur in order to complete a degree. AI models also can assist a financial aid office in targeting limited aid resources to students where the aid offer is most likely to result in a student attending UC. However, similar to the issues listed above in admissions, the potential for

[130] A recent related study showed that algorithms that use multiple factors outperformed the typical placement test when determining whether remedial coursework was required or recommended. This reduced costs for students, improved graduation rates, and corrected for the over-representation of minority students in remedial classes. See Peter Bergman, Elizabeth Kopko, and Julio E. Rodriguez, "Using predictive analytics to track students: Evidence from a seven-college experiment," NBER Working Paper 28948, DOI 10.3386/w28948 (June 2021), https://www.nber.org/papers/w28948.

harm exists if the data used for the algorithm are outdated or based on a limited demographic, or gives a negative decision due to misleading proxies that lead the algorithm to deem the student a credit risk. Concerns arise over accuracy and a lack of transparency in the decision-making of the algorithm, and clearly financial aid decisions have significant potential to affect outcomes for students and applicants.

Although we are unaware of automation used at UC to decide offers of admission, an example comes from the University of Texas, Austin, where the GRADE (GRaduate ADmissions Evaluator) system for graduate admissions was implemented (2013-21). GRADE was a machine learning system developed to reduce the time and effort required of a human admissions committee in reviewing applications within UT Austin's Department of Computer Science (UTCS).[131] GRADE used historical admissions data to predict how likely the committee would be to admit each new applicant.

While it was estimated that GRADE reduced the total time spent in reviews by at least 74%, public outcry over whether the model perpetuated bias and discrimination led to its termination in 2021.[132] GRADE was not used to determine who was admitted or rejected outright; however, critics argued that the model's prediction may lead to biased decisions by the reviewers. Critics questioned whether the choice of features (which emphasized the reputation of the institution students attended previously) could perpetuate existing biases (e.g., in the reviewers' or recommenders' judgment).[133]

Another recent example comes from England, where Ofqual is the regulator of qualifications, exams, and tests.[134] In 2020, the examinations were canceled due to the COVID-19 pandemic. [135] Ofqual then introduced a grades-standardization algorithm that would moderate the teacher-predicted grades for A-level and GCSE qualifications in that year in an effort to prevent grade inflation. That is, the goal was to ensure, as far as possible, that qualification standards were maintained and the distribution of grades followed a similar profile to that in previous years. Teacher rankings were taken into consideration but not the teacher-predicted grades submitted by schools and colleges.

The details of the algorithm were not released until after the results of its first use in August 2020, and then only in part. The important details of the algorithm are as follows. It was based on historical data at the level of each school. It was only used for cohorts with greater than 15 students; otherwise, the teacher-predicted grades were used (which was responsible for the public outcry to a large extent). The algorithm is essentially a form of curve grading, where the curve (grade distribution) was obtained from the school's previous three years. The specific formula is now

---

[131] Ibid.

[132] Austin Waters and Risto Miikkulainen: "GRADE: Machine Learning Support for Graduate Admissions". *AI Magazine* 35(1):64-75 (2014). http://doi.org/10.1609/aimag.v35i1.2504; Burke, Lilah. "The Death and Life of an Admissions Algorithm," *Inside Higher Ed*, December 14, 2020, https://www.insidehighered.com/admissions/article/2020/12/14/u-texas-will-stop-using-controversial-algorithm-evaluate-phd.

[133] Quach, Katyanna. "Uni revealed it killed off its PhD-applicant screening AI -- just as its inventors gave a lecture about the tech," *The Register*, December 8, 2020, https://www.theregister.com/2020/12/08/texas_compsci_phd_ai.

[134] "Ofqual," *GOV.UK*, accessed July 20, 2021, https://www.gov.uk/government/organisations/ofqual.

[135] "Ofqual exam results algorithm," *Wikipedia*, accessed July 18, 2021, https://en.wikipedia.org/wiki/Ofqual_exam_results_algorithm; "2020 UK GCSE and A-Level grading controversy," *Wikipedia*, accessed July 6, 2021, https://en.wikipedia.org/wiki/2020_UK_GCSE_and_A-Level_grading_controversy; Karen Hao, "The UK exam debacle reminds us that algorithms can't fix broken systems," *MIT Technology Review*, August 20, 2020, https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system.

publicly available. Unfortunately, Ofqual could not test the algorithm because they hadn't collected the teacher-predicted grades for previous years. In hindsight, this was a mistake.

The result was that 36% of students received a lower grade from the algorithm, replacing the teacher-predicted one; this disproportionately affected the larger classes often found in state schools, which serve many pupils of a lower socio-economic background. This led to widespread complaints from the public of perceived unfairness, at least, from those who were downgraded. In the end, the grade used was the teacher-predicted grade for A-levels (at 18 years old) and the higher of the teacher-predicted and algorithm grades for GCSE (at 16 years old). This caused top-tier universities to have problems managing capacity, since they received many more above-bar applicants.[136]

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** UC should consider incorporating AI-enabled tools for admissions decisions, if it increases efficiency and offers a counterweight to human bias.

- **Human Values:** Computational models for admissions decisions should incorporate up-to-date considerations used in holistic review. This means that the computational model must be able to take into account difficult-to-quantify criteria such as valuing life experiences as part of a student's capacity for resilience and persistence needed to complete college-level work. If the computational model does not accommodate criteria such as life experience, a human must remain in the loop on that part of the review.

- **Transparency:** Some form of statistical model has long been used at several UC campuses.[137] The models are not public and students are offered no individualized explanation for why they were denied. Transparency about these decisions could be improved but must also be weighed against the possibility for gaming the system (would a sophisticated consultant be able to help an applicant tailor the application to receive a higher score from the model? Such manipulation of perceived preferences also occurs when applications are reviewed by human staff).[138]

- **Fairness/Non-discrimination:** When AI-enabled tools are implemented, users should ensure that their training data are representative of the broad demographic of UC students and applicants. Algorithms can also reinforce biases if, for example, letters of recommendation, which have themselves been shown to demonstrate bias, outweigh other

[136] The EU's AI Regulation has called out the following applications of AI in education and vocational training as high risk: (a) AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions; (b) AI systems for assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions.

[137] Information in this section provided by admissions staff at UC Santa Barbara, UC Riverside, and UCOP.

[138] See the summary and discussion section in Juan E. Gilbert and Andrea E. Johnson, "A study of admissions software for achieving diversity," *Psychology Journal*, Volume 11, Number 1, (2013): 67–90, http://www.psychnology.org/File/PNJ11(1)/PSYCHNOLOGY_JOURNAL_11_1_GILBERT.pdf.

features in the model.[139] On the other hand, AI can provide a check on human biases, helping to increase fairness and accountability in institutional decision-making.[140]

- **Shared Benefit and Prosperity:** AI platforms are to be recommended when, if used well, they can help to correct for human bias and increase equity in admissions decisions and expedite financial aid.

- **Privacy and Security:** Existing University privacy and data security standards and policies should be evaluated to ensure protection of personally identifiable data used within AI applications.[141]

- **Accuracy, Reliability, and Safety:** AI platforms may enhance the accuracy and reliability of decisions regarding admission and financial aid.

- **Accountability:** It is important that admissions directors and staff be held accountable for their decisions, whether they are enhanced or expedited by AI platforms or conducted entirely by humans.

Retention, Student Advising, & Academic Progress

The summer between high school and college can be a precarious time, especially for low-income students and those whose family members have not attended college (i.e., "first-generation" students). They may face decisions and paperwork related to housing, course selection, orientation, and more for which they have little guidance from counselors or others who have navigated this path before them. This is a topic of considerable concern for UC students, some 42% of whom are the first in their families to attend college.[142]

Surprisingly, there is little national data regarding trends in "summer melt" (when a student has been admitted and indicates intent to enroll but fails to register).[143] In 2019, approximately 66% of high school graduates enrolled immediately in a 2- or 4-year college, a rate that has held steady for the last decade.[144] Differences in rates of enrollment appear in the data based on family household income and student grade point averages, factors which also affect the rate of summer melt. These rates may vary from 8% to 40%, depending on the type of institution (higher rates of attrition

---

[139] Iwen, Michelle, "Letters of recommendation: Just say no," *Inside Higher Ed*, April 10, 2019, https://www.insidehighered.com/advice/2019/04/10/letters-recommendation-reaffirm-entrenched-systems-bias-and-exclusion-opinion. Madera JM, Hebl MR, Martin RC, "Gender and letters of recommendation for academia: agentic and communal differences," *Journal of Applied Psychology*, (Nov 2009): 94(6), 1591-9. doi: 10.1037/a0016539. PMID: 19916666.

[140] Jaschik, Scott, "Do college application essays favor wealthier students?" *Inside Higher Ed*, June 1, 2021, https://www.insidehighered.com/admissions/article/2021/06/01/do-college-application-essays-favor-wealthier-students.

[141] Details regarding UC's use and retention practices for data related to undergraduate applicants may be found in this statement: "The University of California Statement of Privacy Practices – General Data Protection Regulation - UCOP Undergraduate Admissions," June 3, 2021, https://apply.universityofcalifornia.edu/docs/StatementOfPrivacy.pdf.

[142] "UC kicks off system wide effort to support first generation students with new report, website," *UC Office of the President*, August 23, 2017, https://www.universityofcalifornia.edu/press-room/uc-kicks-systemwide-effort-support-first-generation-students-new-report-website.

[143] Larry Freedberg, "California's 'cradle-to-career' data system in line to receive $15 million for next phase," *EdSource*, January 12, 2021, https://edsource.org/2021/californias-cradle-to-career-data-system-in-line-to-receive-15-million-for-next-phase/646702.

[144] "Immediate College Enrollment Rate," *National Center for Education Statistics*, May, 2021, https://nces.ed.gov/programs/coe/indicator/cpa.

obtained at 2-year colleges) and family wealth (greater wealth suggests lower attrition rates).[145] For students who have completed college-prep requirements and intend to enroll in a 4-year institution, rates of summer melt in a sample of low-income students are around 8%.[146]

Interventions to decrease the rate of summer melt include peer counseling, professional counseling, and in-person programs, in addition to technological interventions such as texting. Among the few recent studies using treatment and control conditions, such programs show improvements in rates of actual enrollment among those receiving personalized counseling of 3% to 5% and among those receiving interventions via text message of 7%. Costs per student are also significantly less for the technology interventions (about $7 per student) versus the personalized counseling ($80 - $100 per student).[147]

Few studies of texting apps using AI have been conducted by independent researchers. One recent study in eastern North Carolina suggests that AI-enabled chatbots improved student completion of financial aid forms and other administrative tasks in advance of enrollment in a 4-year public university.[148] The effect was most pronounced among first-generation college students. In this study, AI-enabled outreach increased the percentage of students accepting financial aid by 8 points, of registering for classes by 3 points, and enrolling by 3 points.[149] Comparable success was not achieved with a similar intervention planned at a nearby community college. Challenges included lack of staff capacity to implement the platform and incomplete contact information (primarily cell phone numbers) for students.[150]

Texting as a medium of communication with incoming and current students offers a number of advantages. Students tend to engage more easily through texting or instant messaging platforms than via email; in addition, texting can be done on earlier model phones or less expensive devices, not necessarily a smartphone. Texting and chatbots offer the ability to respond to students at scale; especially for institutions with large student populations, it can be unwieldy and expensive to hire a sufficient number of student advisors. Beyond simply broadcasting messages regarding deadlines or upcoming events, AI-enabled chatbots may help students feel they are getting personal attention.

For staff, chatbots may ease the burden on admissions offices by reducing phone and email questions while simultaneously enhancing the student experience and reducing stress. Staff satisfaction could also rise when admissions directors are able to provide more personalized and higher value interactions with first-generation students.

---

[145] Belen Sanchez, "From intended enrollment to actual enrollment: A statistical analysis of summer melt," D. Ed. diss., UCLA, 2020. https://escholarship.org/content/qt5x1210nq/qt5x1210nq_noSplash_aa5a87c7d8c364defdc366cc3ced54a5.pdf?t=qfgnxy.

[146] Many details and variables are discussed in Sanchez's dissertation, based on a sample of 17,434 student records from a national public charter school.

[147] Sanchez, 22-24.

[148] Aizat Nurshatayeva, Lindsay C. Page, Carol C. White, and Hunter Gehlbach. (2020). "Proactive student support using artificially intelligent conversational chatbots: The importance of targeting the technology," (EdWorkingPaper: 20-208). Retrieved from Annenberg Institute at Brown University: https://www.edworkingpapers.com/ai20-20. The research team worked with commercial provider AdmitHub for implementation of the testing platform.

[149] Nurshatayeva, 4.

[150] Sri Ravipati, "Using AI Chatbots to freeze 'summer melt' in higher ed," *Campus Technology*, March 7, 2017, https://campustechnology.com/articles/2017/03/07/using-ai-chatbots-to-freeze-summer-melt-in-higher-ed.aspx; Kelly Field, "This may be the worst season of summer melt in memory. Here's how some colleges are fighting it," *Chronicle of Higher Education*, July 16, 2020, https://www.chronicle.com/article/this-may-be-the-worst-season-of-summer-melt-in-memory-heres-how-some-colleges-are-fighting-it.

Georgia State has pioneered use of chatbots to reduce summer melt. The school introduced an AI-powered chatbot, "Pounce," in summer 2016.[151] In its first summer of implementation, "Pounce delivered more than 200,000 answers to questions asked by incoming freshmen, and the university reduced summer melt by 22%. This translated into an additional 324 students sitting in their seats for the first day of classes." The assistant vice president of undergraduate admissions also noted the cost savings achieved by not having to hire staff: "We would have had to hire 10 full-time staff members to handle that volume of messaging without Pounce."[152]

Retention and degree completion rates are of concern for students and their families as well as the higher ed institutions that serve them. The overall 6-year graduation rate for students enrolled in 4-year bachelor's degree programs in 2012 was 62% (i.e., for students who enrolled in fall 2012, 62% of them had completed their degree by 2018 at the same institution where they started). The rate of degree completion is significantly higher for students attending more selective colleges; at 4-year institutions with acceptance rates less than 25%, the 6-year graduation rate is 90%.[153] Among the UCs only Berkeley and UCLA have acceptance rates below 25%.[154] UC's overall 6-year graduation rate for the cohort entering in 2012 is 84%, more than 20 points higher than the national average.[155] Still, gaps remain in degree attainment, with lower rates for students coming to UC with less preparation, challenging family circumstances, as well as campus climate, availability of courses, and other factors.

AI-enabled platforms and chatbots offer opportunities to increase retention and persistence throughout a 4-year degree. EdSights is one company that offers text-based AI-enabled platforms to improve enrollment and retention.[156] Interventions can help identify students at risk of dropping out or under-performing academically.[157] This is beneficial not only for students but also for the institutions like UC that rely on tuition income to meet financial models.

Such automated platforms may offer advantages: first-generation students may be less self-conscious to ask an AI-enabled chatbot questions than a live person. If the chatbot doesn't know the answer, the question can be forwarded or flagged for a real person. It can also flag language of concern, for example, if students reveal they are sick (e.g., with COVID) or food-insecure. Chatbots can help with administrative tasks, boost mental health, as well as create a sense of community with the school.[158]

Principles Implicated & Sub-Recommendations:

---

[151] "Reduction of Summer Melt," *Georgia State*, accessed July 19, 2021, https://success.gsu.edu/initiatives/reduction-of-summer-melt/.

[152] Ibid.

[153] "Graduation rates," *National Center for Education Statistics*, accessed July 19, 2021, https://nces.ed.gov/fastfacts/display.asp?id=40.

[154] Shirag Shemmassian, *Shemmassian Academic Consulting*, accessed July 19, 2021, https://www.shemmassianconsulting.com/blog/uc-rankings.

[155] "Undergraduate Student Success," *University of California Accountability Report*, accessed July 19, 2021, https://accountability.universityofcalifornia.edu/2019/chapters/chapter-3.html.

[156] "EdSights Retention: Non-Cognitive Data from Your Students Directly," *EdSights*, accessed July 20, 2021, https://www.edsights.io/; Natasha Mascarenhas, "EdSights raises money to help schools reduce their drop-out rates," *TechCrunch*, May 20, 2020, https://techcrunch.com/2020/05/20/edsights-raises-money-to-help-schools-reduce-their-drop-out-rates/.

[157] "How AI chatbots identify at-risk students earlier," *EdSights*, February 8, 2021, https://www.edsights.io/post/why-ai-chatbots-identify-at-risk-students-earlier.

[158] Nina Agrawal, "California College Students Confide in AI Chatbots," *Government Technology*, March 9, 2021, https://www.govtech.com/education/higher-ed/california-college-students-confide-in-ai-chatbots.html.

- **Appropriateness:** AI-enabled tools, such as chatbots, to help students navigate the admissions and first-year process can reduce summer melt and increase retention. However, UC should put in place appropriate oversight processes to ensure these tools do not inadvertently disenfranchise students, especially underrepresented students.

- **Transparency:** Students should be aware of applications that passively collect data about them and be able to turn off certain features that may be overly intrusive (i.e., location data or use of certain campus services). They should be able to opt-out of applications that invite their interaction (such as chatbots).

- **Fairness/Non-discrimination:** AI-powered chatbots have the potential to increase fairness of access to higher ed if they can answer questions posed by a wide range of students. The platforms should be careful about the training datasets to be aware of potential for bias in the answers they provide.

- **Shared Benefit and Prosperity:** AI-enabled tools for student retention are likely to be offered via mobile devices. University administrators should ensure that such tools are device-agnostic and work on the widest range of devices possible, in order not to disadvantage those without the latest technology.

- **Privacy and Security:** Students' privacy should be respected throughout their college careers. The ability to opt-in or opt-out of applications that track their activity will be essential. Some applications may raise questions about the trade-offs of benefits to the student to enhance their academic success versus increased surveillance of their activities.

- **Accountability:** Each UC campus should adopt clear policies regarding the types of data collected and stewardship practices (e.g., deletion) for use in retention, student advising, and academic progress. Students should have the option to request correction and deletion of their data.

Student Mental Health & Wellness

UC has recognized for several years that its students, following national trends, are increasingly experiencing complex mental health concerns and creating demand for mental health services.[159] Accordingly, UC campuses have invested in student mental health and wellness services, ranging from formal, licensed psychological counseling at campus student health centers, to harm-reduction strategies (e.g., PartySafe@Cal, Safe Party @ UCLA), to peer counseling (e.g., UC Irvine's LGBTQ Mentoring Program), and others. Indeed, UC's operating budget request for fiscal year 2019-2020 specifically targeted funding for student mental health services as one of its highest budget priorities, in order to "improve student access to counseling and related services."[160]

Despite UC's desire to support students' mental health and wellness, a 2020 UC report on students' basic needs noted that "many campuses continue to experience a lack of sufficient funding to provide the breadth and volume of services needed to match the growing demand for student

[159] Promoting Student Mental Health: A Guide for Faculty and Staff, *University of California*, 16, https://www.ucop.edu/student-mental-health-resources/_files/pdf/PSMH-guide.pdf.

[160] Budget for Current Operations (2019-2020), *University of California*, 17. https://www.ucop.edu/operating-budget/_files/rbudget/2019-20-budget-summary.pdf.

mental health support."[161] The COVID-19 pandemic has compounded this problem by increasing the mental strain on college students already reporting record levels of psychological challenges.[162] The health crisis highlights the fragility of a system that even before the pandemic was often insufficient to meet students' needs.

Against this backdrop, health providers, their clients, and UC staff may raise questions about whether AI-enabled tools can be effectively and ethically deployed to support student mental health and wellness. In the model of emotionally focused therapy, a counselor or support-provider who is emotionally safe is accessible, responsive, and engaging. These are the qualities experts in the field of AI hope to bring to technology-based mental health care, including those supported by chatbots.

Chatbots are often associated with customer service functions, but these technologies are also being developed to provide psychological services. "Tess" is a chatbot built by clinical psychologists that can interact with users via text. Tess's creators aim to "deliver emotional wellness and coping strategies."[163] Independent research studies have shown that chatting with Tess led to significantly reduced symptoms, on average by -28% for depression and -18% for anxiety.[164] Students may be more willing to disclose mental health and wellness issues to a bot rather than to a human. Tess also enables users to interact in prefered languages. The capacity for treatment in multiple languages could be an advantage for the 25% of California's students who are English-language learners.[165]

Several California State University campuses have implemented a chatbot developed by AdmitHub; the avatar differs by campus (e.g., "Billy" is at Cal Poly Pomona) but they offer similar services. First launched in 2019 to answer basic questions about campus services and deadlines, the chatbots have evolved over the past year to address more mental health and wellness concerns as students grapple with the effects of COVID. The associate vice president of undergraduate studies at Cal State Northridge also noted its role in promoting equity by targeting the needs of first-time and transfer students, many of whom come from low-income or underrepresented communities.[166]

Mental health chatbots are just one example of deployment of AI in connection with mental health. The student data UC already collects or might collect in the future can also be combined with AI-enabled tools to help assess student mental health and wellness and behavioral trends, and target interventions. Dartmouth University's "StudentLife Study," conducted with IRB approval, used

---

[161] The University of California's Next Phase of Improving Student Basic Needs; Regents of the University of California Special Committee on Basic Needs November 2020, *University of California*, 34, https://regents.universityofcalifornia.edu/regmeet/nov20/s1attach.pdf.

[162] Sarah Wood, "Report highlights impact of COVID-19 on college students' mental health," *Diverse Education*, Jan. 13, 2021, https://diverseeducation.com/article/200999/.

[163] "Mental Health Chatbot," *X2*, accessed June 1, 2021, https://www.x2ai.com/.

[164] Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. "Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial." *JMIR Mental Health*, (Dec 2018): 5(4), e64, doi:10.2196/mental.9782, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6315222/.

[165] Laura Hill, "California's English Learner Students," *Public Policy Institute of California*, May 15, 2019, https://www.ppic.org/publication/californias-english-learner-students/.

[166] In addition to Northridge, Cal State campuses that received a grant to develop similar platforms include Pomona, San Marcos, East Bay, Channel Islands, Sonoma State and Humboldt State. Agrawal, Nina, "Artificial Intelligence Meets Real Friendship: Students are Bonding with Chatbots," *Los Angeles Times*, March 9, 2021, https://www.latimes.com/california/story/2021-03-09/college-students-pour-out-emotions-amid-pandemic-to-bots.

passive cell phone data from 48 students over a ten-week period to assess the students' mental health, academic performance, and behavioral trends.[167]

There are many other examples of current or possible future use of AI in connection with student mental health and wellness.[168] And each AI mental health application raises its own ethical challenges, yet there are common elements that bear consideration and drive the recommendations below.

[167] "StudentLife Study," *StudentLife Study*, accessed July 20, 2021, https://studentlife.cs.dartmouth.edu/.

[168] Erica Green, "Surge of Student Suicides Pushes Las Vegas Schools to Reopen," *New York Times*, Jan. 24, 2021, https://www.nytimes.com/2021/01/24/us/politics/student-suicides-nevada-coronavirus.html; Dekker, Izaak et al., "Optimizing Students' Mental Health and Academic Performance: AI-Enhanced Life Crafting," *Frontiers in Psychology* 11 (June 2020), https://www.frontiersin.org/article/10.3389/fpsyg.2020.01063; Jessica Kent, "What Role Could Artificial Intelligence Play in Mental Healthcare?" *Healthcare Analytics*, April 23, 2021, https://healthitanalytics.com/features/what-role-could-artificial-intelligence-play-in-mental-healthcare.

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** UC should encourage student representation, engagement, and feedback in consideration of the appropriateness of an AI-enabled intervention in the context of student mental health and wellness.

- **Transparency:** Any use of AI-enabled tools to address student mental health and wellness, especially passive data collection, should be done with the full consent of the student. Policies regarding data collection and retention should be clear to users.

- **Fairness/Non-discrimination:** Datasets used to inform AI-enabled mental health services should come from a range of student experiences and backgrounds. The algorithms should be adaptive to the needs and experiences represented across student subcultures.

- **Shared Benefit and Prosperity:** Chatbots or other AI-enabled platforms could democratize access to mental health care by reducing costs and enhancing convenience for students who may be juggling school work with employment or family obligations. If considering implementing a chatbot to address student mental health, providers should ensure it will serve the broadest campus population, including non-native English speakers.

- **Privacy and Security:** As with any medically sensitive data, the highest levels of privacy and security should be ensured for data related to mental health.

- **Accuracy, reliability, and safety:** An AI-enabled chatbot should strive for accuracy when evaluating mental health risks and respond reliably to variations in vernacular or current texting abbreviations. It must have the capability to escalate for intervention by a human or a licensed professional.

- **Accountability:** Both the commercial provider and the University should be held accountable to review the performance of the AI-enabled intervention and evaluate efficacy and any potential adverse outcomes.

Grading & Remote Proctoring

In light of resource challenges, higher education institutions are exploring the efficiency, effectiveness, and equity of AI-enabled tools for grading and remote proctoring.[169] Higher education institutions have deployed AI-enabled virtual teaching assistants. The Georgia Institute of Technology uses "Jill Watson," which Georgia Tech describes as "a graduate-level teaching assistant who can hold office hours 24/7/365."[170] Jill Watson can answer a significant proportion (up to 40%) of the many questions students ask each semester. In the era of COVID when all instruction and evaluation took place remotely, new tools were introduced for remote proctoring. Not all of them involve AI; some took advantage of a combination of Zoom video monitoring and restricting access to other websites or devices aside from those on which students were taking an exam. Others have involved automation of monitoring, including machine learning and facial

[169] Expertise to address these questions may be found in offices of Educational Technology Services on a given campus. "Managing Logistics in Large-Enrollment Courses (Dialogues Recap)," *Managing Logistics in Large-Enrollment Courses (Dialogues Recap) | Academic Innovation Studio*, November 29, 2017, https://ais.berkeley.edu/news/managing-logistics-large-enrollment-courses-dialogues-recap.
[170] "Jill Watson," *Georgia Tech College of Computing*, December 14, 2017, https://www.cc.gatech.edu/holiday/jill-watson.

recognition technology. These have introduced opportunities for bias and discrimination when the platforms do not recognize students of color, especially women of color, or have discriminated against Asian students whose facial features differ from typical subjects found in image training sets.[171] In response, prominent advocacy groups issued a letter urging school administrators to ban use of remote proctoring technologies.[172]

Most UC campuses have experimented with remote proctoring tools for test-taking, especially during the academic year in which COVID prevented many in-person classes (2020-21).[173] In response to ethical concerns about the technology's use, UC Berkeley developed a white paper in December 2020, "Remote Exam Proctoring Policy Recommendation."[174] The report emphasizes that new tools must go through an approval process in accordance with campus contracting and purchasing policies (typically via Supply Chain Management) and be carefully reviewed for accessibility requirements and privacy protections.[175]

Use of remote proctoring technologies have increased across numerous UC campuses. Berkeley's Haas School of Business (Haas) uses a remote proctoring tool called Honorlock.[176] Haas encourages instructors to let students know early in the course and clearly state on the syllabus that the platform will be used for exams. Honorlock's website offers details regarding the kind of information collected from the student (including browsing activity and face detection) during the exam. UC Irvine, UC Davis and UCLA offer a combination of remote monitoring and locking down the browser during online testing through a tool called "Respondus" on the Canvas online learning platform.[177] UC San Diego uses "ProctorU" for remote exam proctoring.[178]

Grading apps have been developed in recent years, including at UC. Four UC Berkeley researchers developed a program to help grade papers during their time working as teaching assistants. The team launched the online grading app Gradescope in 2014 and since then the program has graded 250 million questions from a wide range of college courses. AI features of Gradescope address three challenges: identifying question types, distinguishing between different written marks (formulas, sketches, diagrams), and recognizing handwriting. AI helps turn grading into an automated, highly

---

[171] Avi Asher-Schapiro, "Online exams raise concerns of racial bias in facial recognition," *Christian Science Monitor*, Nov. 17, 2020, https://www.csmonitor.com/Technology/2020/1117/Online-exams-raise-concerns-of-racial-bias-in-facial-recognition; David Leslie, "Understanding bias in facial recognition technologies: an explainer," *The Alan Turing Institute*, (September 2020), https://doi.org/10.5281/zenodo.4050457.

[172] Rebecca Klar, "Advocacy groups urge school administrators to ban proctoring," *The Hill*, July 08, 2021, https://thehill.com/policy/technology/562152-advocacy-groups-urge-school-administrators-to-ban-eproctoring.

[173] These examples are a select list. UCOP offers a more comprehensive list of remote assessment and proctoring tools used across the system at https://www.ucop.edu/educational-innovations-services/covid19/remote-assessment-and-proctoring.html.

[174] "Remote Exam Proctoring Policy Recommendation," *Remote Exams Working Group*, December 11, 2020, https://drive.google.com/file/d/19Agbh3NvQGfx0jJ8H2koK0TlGFdrPltX/view

[175] Ibid, p. 2.

[176] "Remote Proctoring," *Berkeley Haas*, December 10, 2020, https://haas.berkeley.edu/haas-digital/technology-tools-for-online-teaching/remote-proctoring/; "Online Exam Proctoring with a Human Touch," *Honorlock Proctoring*, accessed July 20, 2021, https://honorlock.com/.

[177] Respondus Monitor uses students' webcams with AI analysis: "Respondus Monitor is a fully-automated proctoring solution. Student's use a webcam to record themselves during an online exam. Afterward, flagged events and proctoring results are available to the instructor for further review." Quote from the company website: https://web.respondus.com/he/monitor/. See details for UCI: https://sites.uci.edu/teachanywhere/home/assessment/respondus/ and UCLA: https://humtech.ucla.edu/instructional-support/using-ccle-assessment/using-respondus/

[178] "ProctorU UCSD," *ProctorU Portal | UCSD*, accessed July 20, 2021, https://www.proctoru.com/portal/ucsd.

repeatable exercise by learning to identify and group answers, and thus treat them as batches.[179] AI-enabled tools also can address academic dishonesty in written work, such as the platform Turnitin, which can detect plagiarism using "machine intelligence."[180] (Turnitin acquired Gradescope in 2018.[181])

Principles Implicated & Sub-Recommendations:

- **Appropriateness:** Instructors should consider whether use of online tools is warranted for a given class. The ability to grade or proctor work for a large-enrollment class may prompt use of AI applications more than for smaller classes. The kind of testing (e.g., essay response versus discrete answers) may also be relevant to decisions over appropriate use.

- **Transparency:** University instructors must alert students when automated grading or proctoring tools will be used for evaluation. The University affords students various processes to challenge or appeal grading decisions or allegations of academic misconduct. Some of the processes reflect legal requirements to provide adequate due process where the University proposes disciplinary action. An AI-enabled tool that is not explainable and easily understood by students and instructors would undermine the University's grade appeal and student conduct processes because a student or instructor would not be able to understand how the instructor or University reached the decision and, therefore, would not be able to challenge the decision effectively.

- **Fairness/Non-discrimination:** Use of AI-enabled tools for grading and proctoring must be carefully vetted to avoid discrimination and bias. As noted above, features such as facial recognition have well-documented risks related to racial bias. Other features could "flag" as suspicious benign behavior that relates to disabilities (e.g., keystroke analysis). Gendered teaching bots may also reinforce gender stereotypes.[182]

- **Privacy and Security:** Instructors should weigh the need to ensure academic integrity against potential privacy and cybersecurity concerns. They should also ensure secure data practices (cybersecurity risk management, data storage, management, and deletion) for the third-party vendors they may use.

- **Accuracy, Reliability, and Safety:** Accuracy, reliability, and security are key concerns for AI-driven evaluation and grading tools. UC should ensure appropriate oversight mechanisms, including data governance strategies, are in place to continuously evaluate performance and ensure data privacy and security.

- **Accountability:** Students should have an opportunity to challenge outcomes decided by AI-enabled grading or evaluation tools. Teachers and students must hold the vendors accountable for anomalous outcomes (i.e., certain demographics of students being "flagged" for review more frequently than others).

---

[179] Kirsten Mickelwait, "Gradescope: Taking the Pain out of Grading," *Berkeley Engineering News*, January 15, 2016, https://engineering.berkeley.edu/news/2016/01/gradescope-taking-the-pain-out-of-grading/.

[180] "About Us: About Turnitin, Our Mission & Values," *Turnitin*, July 20, 2021, https://www.turnitin.com/about.

[181] "Terms of Use | Gradescope," *Gradescope*, July 20, 2021, https://www.gradescope.com/tos.

[182] Caitlin Chin and Mishaela Robison, "How AI Bots and Voice Assistants Reinforce Gender Bias," *Brookings*, November 23, 2020, https://www.brookings.edu/research/how-ai-bots-and-voice-assistants-reinforce-gender-bias/.

Recommendations on Implementation of UC AI Principles in Student Experience

Use of an AI-enabled tool may be preferable to the status quo if it introduces efficiencies and reduces (or exposes) human bias. Both the benefits and the harms of AI should be considered in a specific application and as part of a larger system. We are still in the early stages of widespread adoption of AI, and some hazards are poorly understood both from a technical standpoint (e.g., the algorithm) and in terms of ethics and governance. Based on the above considerations, we recommend that UC engage in the following next steps:

1. Develop a vetting process for AI-enabled tools that affect students.

2. Put in place appropriate safeguards to mitigate risks of algorithmic bias.

3. Educate stakeholders on the risks of AI-enabled tools that affect students.

4. Develop a database that inventories and tracks use of AI-enabled tools used in the "student experience" activities outlined in this section (e.g., in exam proctoring and grading) and put in place transparency procedures.

5. Implement appropriate privacy-preserving methods in the development and use of AI-enabled tools for students.

Each of these recommendations is explored in greater detail below.

Vetting Process for AI-enabled Tools

Students should be involved in the decision-making process for deploying AI-enabled tools within the scope of the "student experience" topics addressed in this section. AI-enabled tools for any student-facing applications should be assessed for bias and discrimination. Specifically, UC should adopt tools in this area only where sufficient information is available about datasets used to train the models and on the actual track record of the tool in practice to ensure it will not differentially affect students based on race, disability, or other protected categories. Those seeking to develop and/or implement AI-enabled tools in student experience should engage the University's equity and inclusion experts in evaluation of the tools during development and adoption and at regular intervals during implementation.

Application of AI-enabled tools in the area of mental health should have additional oversight. If these tools are implemented for mental health applications, it is essential that clinicians are included in the vetting process and educated on how to use AI technology with their patients. If AI becomes a regular part of clinical practice (particularly to perform documentation), more information is needed on the technology's therapeutic benefits—especially compared to the potential risks. Qualified mental health professionals or academics should review the application and provide guidance on efficacy and safety considerations (e.g., what keywords or behaviors might trigger human review). AI should not replace clinicians in diagnosing conditions or prescribing medical or psychiatric treatment.

Mitigating Risk of Algorithmic Bias

In order to mitigate risks of algorithmic bias, documentation of how the AI-enabled tool was developed should be provided, including how the data was collected and used and what features were selected, such as sample size, population sampled from, and time of collection. Information on how the model was trained on data collected, including information such as loss function, regularization, training algorithm, and model selection, should also be provided.

The training dataset and mode of training should be carefully evaluated for bias before deployment. When using an AI-enabled tool with human data, such as in admissions decisions, there is a risk that the model may be biased (e.g., discriminate against some populations). From a research perspective, the potential for bias within a model is at present not clearly understood, agreed upon, or even well defined. It is clear that the major contributing factors for bias come from the choice of training data and model features. This follows from the fact that a machine learning model primarily tries to replicate predictions and patterns found in the training data (and generalize to unseen data). The training dataset and how the model is trained (e.g., choice of regularization) may perpetuate biases.

Educating and Involving Stakeholders on AI

Although any use of AI could incur risks, the perceived risk among the general public may also be overly heightened due to sensationalist media articles. Such coverage may lead to an automatic negative perception of "AI systems" or the "algorithm." In particular for college or job applications, this includes misgivings about fairness, bias, and accountability. Because of the relative novelty of AI as a practical technology, many UC employees (at different levels) may not be aware of what constitutes AI, whether a vendor's product may use AI, or the potential advantages and disadvantages of its use.

This suggests the need to:

- Establish a basic form of education for those employees who have decision power (e.g., in purchasing) or who would use AI.

- Educate stakeholders about how AI works, and ideally using transparent AI models. This is particularly important given that making a model public and/or interpretable may not be sufficient to create trust. Indeed, this was true (at least partly) with the recent cases of GRADE and Ofqual (both admissions AI systems), where the model was interpretable and publicly available (although somewhat delayed in the Ofqual case).

- Involve IT staff and procurement specialists in the development of checklists or guidelines for vetting products that apply AI. Policies should be clear at the campus level that departments and schools must receive approval for implementing AI-driven tools for student-facing applications.

Development of a Database and Transparency Procedures

Students should be informed when and how AI-enabled tools are used within the scope of the "student experience" topics addressed in this section. To support this goal, campuses should develop a database that inventories and tracks AI-enabled tools used. UC must be prepared to address audits of AI-enabled tools we use, or public information requests (which may include training data and models). This would be greatly facilitated if using transparent, publicly available

models, so it is clear what data is used and how a model is used. UC should inform users and affected stakeholders of the extent to which a human is “in the loop” on decision-making that uses AI. UC should also clarify and clearly communicate data collection and retention policies for all student-related data used in AI-enabled tools.

Privacy

The use of AI models requires personal data, either to apply the model (e.g., to predict admission) or to train the model (historical admissions results). While use of personal data implicates privacy concerns, the use of AI may sometimes help in preserving privacy. For example, in remote proctoring via a video camera, video is sent to a human proctor, which is intrusive. Instead, it may be possible to have the AI-enabled tool run locally (or send only "video fingerprints" to a server), so that no video is actually sent unless strictly necessary (e.g., if the AI detects a suspicious pattern). Privacy should be kept in mind (and University privacy officials consulted) whether the system is built in-house or bought from industry.

CONCLUSION: RECOMMENDATIONS TO GUIDE UC'S AI STRATEGY

Use of AI may enable UC to increase the efficiency, effectiveness, and equity of its operations. However, ill-considered applications may result in detrimental outcomes by perpetuating existing failings in current processes, creating new vulnerabilities, or ingraining biased and discriminatory practices. While the Working Group recommends that each campus be empowered to determine how to appropriately implement the following recommendations—taking into account each campus's needs and values and conformity with campus-specific legal and policy compliance—we strongly believe efforts should be coordinated across UC campuses to reduce redundancies and inefficiencies and support the collective adoption of best practices.

Drawing upon the insights and recommendations from the Working Group subcommittees, we provide the following four primary recommendations to guide UC's development and use of AI in its operations.

1. Institutionalize the UC Responsible AI Principles in procurement, development, implementation, and monitoring practices

AI-specific governance and oversight processes should be established to inform UC's procurement, development, implementation, and monitoring of AI used in its operations. In doing so, the shared principles help to focus and guide a coordinated strategy across UC. As a first step, the UC Responsible AI Principles proposed in this report should be institutionalized within UC-wide procurement policies and practices. One potential strategy proposed in the federal government is to establish a training program for procurement officers to learn about AI, risks posed, and risk mitigation strategies.[183] We recommend that a similar training program be established for UC procurement officers, but should also include guidance on effective strategies to operationalize the UC Responsible AI Principles in procurement policies and related oversight processes. The campus-level councils (discussed next) can serve as an educational resource and assist in the development of a training program.

Training and implementation of oversight processes during procurement, development, implementation, and monitoring of AI-enabled tools deployed in UC services are critical to identifying and mitigating risks before widespread harm. For implementations that do not go through a formal procurement process, faculty and staff implementing the AI-enabled tool should disclose proposed and/or actual use of the AI-enabled tool to their campus-level council (detailed below). If the tool is applied in an area that poses greater than moderate risk to individual rights, its use should also be documented in a public AI database (detailed below), including potential risk(s) posed by the use of the tool and risk mitigation strategies implemented.

2. Establish campus-level councils and support coordination across UC that will further the principles and guidance developed by this Working Group

To help guide appropriate procurement, development, implementation, and monitoring of AI in alignment with the UC Responsible AI Principles, multistakeholder campus-level councils should be established. The councils should include diverse faculty, staff, and student representatives to better ensure appropriate operationalization of the UC Responsible AI Principles and that the diverse

[183] Dave Nyczepir. "Senators Propose AI Training for Federal Acquisition Workforce," *FedScoop*, Aug. 9, 2021, https://www.fedscoop.com/ai-training-acquisition-workforce-bill/.

viewpoints, needs, and priorities of the UC community are considered and addressed in the implementation of AI within university services. The councils should also serve as an educational resource, including identifying ways to help educate campus decision-makers on the benefits and risks of AI and appropriate risk-mitigation strategies. Furthermore, the councils should ensure that each campus develops appropriate data stewardship standards for UC data that may be used in the development and use of AI-enabled tools and systems (e.g., Require procurement and contract offices to utilize the UC Responsible AI Principles to evaluate all agreements that involve sharing UC data with third-parties.)

To help facilitate lesson sharing and coordination of best practices within and across all UC campuses, the councils should draw upon the expertise and personnel of established councils, committees, working groups, etc. already addressing related issues on their campus. For example, most UC campuses have established bodies dealing with information risk management and data privacy and security, such as Information Risk Governance Committees, Campus Information Security and Privacy Committees, and Data Governance Boards. To support coordination across campuses, the campus-level councils should each elect one to two representatives to participate in regular joint meetings of the campus-level council representatives with UCOP personnel.

3. Develop an AI risk and impact assessment strategy

To support appropriate evaluation and oversight of AI, UC should develop a shared risk and impact assessment framework and implementation strategy that can be used to help identify and evaluate risks of AI-enabled tools and appropriate risk mitigation strategies. Development of the risk and impact assessment framework and corresponding strategy should be a multistakeholder process, including members of the campus-level councils and other faculty, students, and staff representatives from UC Legal; Office of Ethics, Compliance and Audit Services (ECAS); Procurement; Office of the Chief Information Officer; Research Policy Analysis and Coordination (RPAC), among others.

AI risk and impact assessments are gaining prominence as viable oversight strategies, increasingly being incorporated into legislation and regulation.[184] Lessons learned from implementations in the public and private sectors can help to guide the UC's strategy.[185] While the specific criteria outlined in established AI risk and impact assessments vary, each typically evaluates the nature, scale, duration, irreversibility, and likelihood of potential negative effects. Risk mitigation strategies are then calibrated in response to the collective assessment. Most established AI risk and impact assessments utilize tiered risk levels (e.g., low, moderate, high, and existential risk) posed to individual rights. Applications posing greater than moderate risk are required to go through a thorough review and risk mitigation process. We recommend that the campus-level councils collaboratively assist in the development of the AI risk and impact assessment strategy, including determining risk levels and the thresholds that will trigger oversight processes.

---

[184] The House Appropriations Committee has directed NIST to develop an "AI risk framework" to guide the "reliability, robustness, and trustworthiness of AI systems" in the federal government. "Commerce, Justice, Science And Related Agencies Appropriations Bill, 2021 - Report Together With Minority Views," *House Committee on Appropriations*, July 2020, https://appropriations.house.gov/sites/democrats.appropriations.house.gov/files/July%209th%20report%20for%20circulation_0.pdf, 23.; The European Commission's "AI Act" calls for the implementation of "conformity assessments" to evaluate and mitigate risks of AI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

[185] See the Ethics and Algorithms Toolkit: https://ethicstoolkit.ai/' and Canada's Algorithmic Impact Assessment tool: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.

4. Document AI-enabled technologies in a public database

To support greater transparency and accountability, AI-enabled technologies implemented within UC services that pose greater than moderate risk to individual rights (e.g., in HR; health; policing; and student services, such as remote proctoring and grading) should be recorded in a public database. We recommend that each campus establish its own database. To the extent practicable, the databases should have the same structure and data points to enable cross-campus insights and comparisons.

The campus-level councils should collaboratively determine what criteria should be documented, such as data, models, potential risks, and risk mitigation strategies. We recommend the following criteria be considered for inclusion in the database: description of the AI-enabled tool and who developed it; intended uses, limitations, and risks; risk mitigation strategies; when possible, the algorithms and training data used; and information on appropriate channels individuals can pursue to provide feedback or contest a decision made by an AI-enabled tool. Documentation strategies implemented in the private sector can provide guidance, such as IBM's "FactSheets" that encourage AI developers to record the "purpose, performance, safety, security, and provenance" of AI models to build trust and accountability.[186] Google's "Model Cards" provide a framework for recording the "provenance, usage, and ethics-informed evaluation" of AI models and Microsoft's "Datasheets for Datasets" and "Transparency Notes" encourage documentation of data provenance, composition, collection processes, motivations, and recommended uses.[187] UC should encourage public documentation of AI models and data used. In doing so, UC is taking important steps toward greater transparency and accountability.

[186] Matthew Arnold et al., "FactSheets: Increasing trust in AI services through supplier's declarations of conformity," *IBM Journal of Research and Development* 63, no. 4/5 (2019): 6-1.; Microsoft, "Transparency Note for Text Analytics", Dec. 4, 2020, https://docs.microsoft.com/en-us/legal/cognitive-services/text-analytics/transparency-note.

[187] Huanming Fang and Hui Miao, "Introducing the Model Card Toolkit for Easier Model Transparency Reporting," *Google AI Blog* (blog), *Google*, July 29, 2020, https://ai.googleblog.com/2020/07/introducing-model-card-toolkit-for.html; Timnit Gebru et al., "Datasheets for datasets." *arXiv preprint arXiv:1803.09010* (2018).

ACKNOWLEDGMENTS

We thank the members of the UC Presidential Working Group on AI for their contributions to this report. We are especially grateful to Shanda Hunt, Systemwide Research Compliance Officer at UCOP for her significant contributions to the Working Group, without which we would have not been able to complete our work. We are also grateful for the guidance we received from campus Chief Information Officers and Chief Technology Officers.

We would like to thank the following individuals for their helpful guidance and support:
Jennifer Lofthus, General Compliance Manager, Office of Ethics, Compliance & Audit Services, UCOP
Pegah Parsi, Campus Privacy Officer, UC San Diego
Scott Seaborn, Campus Privacy Officer, UC Berkeley
Justin Sullivan, Executive Director, Strategic Sourcing at the University of California
Noelle Vidal, Healthcare Compliance and Privacy Officer, UCOP
Kent Wada, Chief Privacy Officer, UCLA

HEALTH
Jason Y. Adams, Associate Professor, Division of Pulmonary, Critical Care, and Sleep Medicine, UC Davis School of Medicine and Director of Digital Health Innovation, UC Davis Health
Mary Alexander, Research Compliance Officer, UCI Health
Tom Andriola, Vice Chancellor, Information, Technology and Data; Chief Digital Officer, UC Irvine and UC Irvine Health
Atul Butte, Chief Data Scientist at UC Health and Director of the Bakar Computational Health Sciences Institute at UC San Francisco
Rachael Callcut, Vice Chair Clinical Science and Division Chief, Trauma, Acute Care Surgery, and Surgical Critical Care, UC Davis Health
Albert Duntugan, Chief Data Officer, UCLA Health
Mike Hogarth, Professor, Division of Biomedical Informatics, Department of Medicine, and Clinical Research Information Officer, UCSD Health
Chris Kello, Interim Vice Provost and Graduate Dean, Professor of Cognitive and Information Sciences, UC Merced
Shamim Nemati, Director, Predictive Health Analytics and Assistant Professor, UCSDH, Department of Biomedical Informatics
Ziad Obermeyer, Associate Professor and Blue Cross of California Distinguished Professor of Health Policy and Management at the School of Public Health, UC Berkeley
Naveen Raja, Medical Director of Population Health and Associate Clinical Professor, David Geffen School of Medicine at UCLA
Vanessa Ridley, Chief Compliance Officer, UCSF
Laurel Riek, Associate Professor of Computer Science and Engineering and Director, Healthcare Robotics Lab, UC San Diego

HR
Genevieve Smith, Associate Director, Center for Equity, Gender & Leadership, Haas School of Business, UC Berkeley

POLICING
Roberto Meza, Campus Physical Security Program Manager, UC San Diego Police Department

STUDENT EXPERIENCE
Emily D. Engelschall, Director of Admissions, UC Riverside
Lisa Przekop, Director of Admissions, UC Santa Barbara
Michelle Whittingham, Associate Vice Chancellor for Enrollment Management, UC Santa Cruz
Han Mi Yoon-Wu, Executive Director, Systemwide Undergraduate Admissions, University of California

APPENDIX: RELEVANT LAWS, REGULATIONS, & POLICIES

This appendix highlights relevant laws, regulations, and policies to consider in connection with the procurement, development, implementation, and monitoring of AI used in UC services. It is not comprehensive of *all* relevant laws, regulations, and policies and corresponding legal issues, as that is beyond the scope of this report. This appendix should not be construed as legal advice. Legal counsel should be consulted on each campus.

While framed as privacy laws, at its core, the California Consumer Privacy Act (CCPA) and its amendment, the California Privacy Rights Act of 2020 (CPRA 2020) are intended to address the right for an individual to control how data about them is used.[188] These laws set forth requirements for businesses that collect, share, or sell personal information about California residents. Specifically, it gives consumers: (1) the right to know what personal information about them is being collected and how it is used and shared; (2) the right to control how their personal information is used; and (3) the right to equal service and price, even if they exercise their privacy rights. These laws directly apply to for-profit businesses, and therefore will not directly apply to UC. However, they will apply to UC where it receives data from a business subject to the laws (e.g., when it engages in a service with the business or receives the data for research purposes). In such cases, UC may be contractually obligated to assist the business in ensuring that California residents can exercise their rights under the laws, including, but not limited to, providing information about how their data is used or even deleting data.

The laws are modeled after the European Union's General Data Protection Regulation (GDPR). The GDPR applies to any individual located in the European Economic Area.[189] Notably, GDPR also regulates automated processing.[190] GDPR gives data subjects the right not to be subject to a decision based solely on automated processing, including profiling, where there is a legal or similarly significant effect on the person.[191] GDPR only allows decisions to be made based on automated processing where the individual explicitly consents to the activity and where the controller of the data implements measures to safeguard the person's data and rights under GDPR. The safeguards include the right of the individual to be provided meaningful information about the logic involved, the potential consequences to the data subject, as well as the right of the data subject to obtain human involvement in the processing, and to challenge the decision.[192] Unlike California's laws, GDPR does directly apply to UC when it offers goods or services to individuals living in Europe or monitors the behavior of individuals living in Europe. Thus, it could apply to UC's use of data in AI-enabled tools.

---

[188] The CCPA went into effect on January 1, 2020. CPRA 2020 was passed by California voters pursuant to Proposition 24 on the November 2020 ballot. It expands rights under CCPA, CPRA 2020 is effective January 1, 2023. The New Privacy Laws are codified at Cal. Civ. Code §§ 1798.110-1798.199.95 (West 2020).

[189] The GDPR applies to natural persons "who are in the Union," and not merely citizens or residents. Article 3. The European Economic Area includes the European Union, as well as Iceland, Liechtenstein and Norway. The United Kingdom has enacted a law similar to GDPR upon its exit from the EU. Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation ("GDPR")).

[190] Automated processing is not defined by GDPR, but guidelines on automated individual decision-making and profiling provide that it is the ability to make decisions by technological means, using any data, whether personal data or not. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

[191] GDPR, Art. 22(1).

[192] GDPR, Art. 22.

We next provide a summary of relevant laws, regulations, and policies affecting UC's procurement, development, implementation, and monitoring of AI within the application areas of health, HR, policing, and student experience. Again, the laws, regulations, and policies discussed below are not comprehensive and legal counsel should be consulted on each campus.

Laws & Regulations Affecting AI in Health

While there is no comprehensive federal or state law that regulates AI in health, there are a number of laws that apply to various uses of AI in the health domain. Indeed, the legal framework for the regulation of AI, particularly in health, borrows from a cross-application of rules and regulations from data privacy, discrimination, product regulation, and common law principles. This section provides a brief overview of the most relevant laws and their relevance to UC at the current time, focusing on privacy and security, product safety, and transparency in business practices and use of data. As the discussion below demonstrates, while the legal framework is important to understand, the patchwork and gaps revealed highlight the importance of the work this Working Group is doing to develop overarching ethical principles governing the current and future use of AI.

Privacy and Security

In the United States and California, a number of sector-specific laws protect an individual's health data, largely based on the actor maintaining the information. Under federal law, the Health Insurance Portability and Accountability Act (HIPAA) sets forth privacy and security requirements that covered entities and their business associates must follow.[193]

UCH hospitals and providers are covered entities under HIPAA; thus, where AI is used by UC's hospitals or providers, HIPAA applies. Similarly, when a UC medical center contracts with a third party, such as a software service providing AI-related products to UC's medical centers, HIPAA governs not only UC, but the third party as a business associate.[194]

California's health privacy law, the Confidentiality of Medical Information Act (CMIA), similarly imposes guardrails upon the disclosure of health-related information. However, CMIA only applies to UC's healthcare providers, and only to UC's campuses and researchers to the extent they maintain health records of employees or students. CMIA obligations do not extend to third parties.[195]

Unfair or Deceptive Business Practices: FTC Act, Unfair Business Competition Laws

The Federal Trade Commission (FTC) primary enforcement statute is Section 5 of the FTC Act, which prohibits unfair or deceptive acts or practices in or affecting commerce.[196] The FTC recently

---

[193] See definition of "covered entity" at 45 C.F.R. § 160.103 (2014).

[194] 45 C.F.R. §§ 164.502(e), 164.504(e) (2013).

[195] Confidentiality of Medical Information Act, Cal. Civ. Code §§ 56-56.37 (2020); "Joint Guidance on the Application of the Family Educational Rights and Privacy Act (FERPA) And the Health Insurance Portability and Accountability Act of 1996 (HIPAA) to Student Health Records", U.S. Department Health & Human Services (Dec. 2019), https://studentprivacy.ed.gov/sites/default/files/resource_document/file/2019%20HIPAA%20FERPA%20Joint%20Guidance%20508.pdf.

[196] 15 U.S.C. § 45(a)(2) (2006).

issued guidance indicating that enforcement of Section 5 could include the sale or use of racially biased algorithms.[197] The FTC does not have enforcement jurisdiction over the University of California, as a non-profit organization, but their enforcement authority would extend to companies that UC might contract with to obtain AI-enabled tools.

Common law principles would apply to UC's use of AI technology in health. For example, healthcare providers and UC medical centers may be deemed to owe a duty of care, as well as a fiduciary duty to patients, which could subject UC to damages for negligence or breach of fiduciary duty.

Product Safety: FDA Regulation

The U.S. Food and Drug Administration (FDA) regulates medical devices that use AI-enabled software, including those developed by or used by UC. The FDA proposed a regulatory framework for modifications to AI/ML-based software as a medical device (SaMD) without the need for additional FDA review.[198] Most recently, the FDA issued an action plan that calls for increased transparency in the development and performance of SaMD, methodologies to identify and improve the fairness and robustness of ML algorithms, and the ability for these devices to utilize real-world performance monitoring to respond proactively to safety or usability concerns.[199]

Genetic Data - GINA

The Genetic Information Nondiscrimination Act (GINA) protects health insurers from engaging in genetic discrimination and prevents employers from obtaining and using genetic information in employment decisions. GINA applies to UC as both an insurer and employer.[200]

While there are myriad laws that might impact UC's use of AI-enabled tools in the health domain, they exist in a patchwork, leaving gaps in protection. For example, a UC-developed AI-enabled tool used for administrative or clinical purposes would not be covered by the FTC Act or CCPA. Nor, to the extent this tool did not constitute a device headed towards the market, would it be regulated by the FDA. Conversely, a company that sold an AI-enabled tool to UC would likely be covered by the FTC Act and potentially by other statutes as well but may not have to comply with the HIPAA rules or CMIA. In both instances there would be inconsistent requirements around the treatment of health data.[201]

---

[197] "Aiming for truth, fairness, and equity in your company's use of AI," April 19, 2021, *Federal Trade Commission*, https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai

[198] "Artificial Intelligence and Machine Learning in Software as a Medical Device," *U.S. Food & Drug Admin.* https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.

[199] "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan," January 2021, *U.S. Food & Drug Admin.* https://www.fda.gov/media/145022/download.

[200] Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (codified as amended in scattered sections of 29 & 42 U.S.C.); *see also* Discrimination: genetic information, S.B. 559, 2011-2012 Reg. Sess. (Cal. 2011), https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201120120SB559

[201] The subcommittee notes that the California legislature has been increasingly active in proposing legislation that might impact UC's use of AI-enabled tools and/or health data. To give one example for the 2020-2021 legislative season, proposed AB-13 would require oversight for public contracts that include automated decision systems, including in the area of health. If passed, partnerships between UC and the California Department of Public Health would need to ensure such oversight if the partnership results in the government procuring health-related automated decision systems. See AB-13 Public contracts: automated decision systems, AB-13 2021-2022 Reg. Sess. (Cal. 2021), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB13

Finally, although medical ethics in the US has been informed by guiding principles including autonomy, beneficence, nonmaleficence, and justice for decades, and these remain valid irrespective of the technologies used, these principles—drafted long ago—are insufficient to address all of the challenges presented by the use of AI in the health domain.[202] Accordingly, a more consistent approach based on the AI Ethical Principles is needed.

Existing Oversight

In recognition of the gaps described above, there are ongoing efforts at UCH to provide oversight over UCH data use and sharing, including a few specifically directed at AI-enabled tools that are relevant to and consistent with the work of the health subcommittee. For example, a group at UCSD developed an ethical framework for certain AI health applications, called the Digital Health Checklist.[203] In addition, UCSD Health has a process to review cost, efficiency, user experience, workflow, purpose, accuracy, impact, equity and unintended discrimination/racism in AI-enabled tools.[204] UCLA Health has a development cycle plan to review and measure AI-enabled tools/systems after deployment.[205] Some of the factors that are evaluated include risk, accuracy, unintended consequences, bias, workflow, and outcomes.

In the research context, the academic health campuses rely on Institutional Review Boards (IRB) to review proposals, but IRB review takes place at each campus so there is variation in how ethical issues are managed, including with AI applications. Campuses have expressed interest in having overarching and domain-specific AI guidelines from UCOP to guide the development and application of AI in ways that are consistent with UC's mission and values.

There are also data governance practices in place to promote the safe and responsible use and sharing of UC health data. A report to the President in January 2018 from a UC task force on health data governance identified core principles and recommendations relating to sharing UC Health data with parties outside of UC.[206] The report highlighted gaps in the existing regulatory and legal frameworks for health data and, among other things, called for a principle of data justice, rather than merely focusing on data privacy as a model for health data use at UC. The focus on justice includes enabling community input to shape scientific and health goals, creating novel strategies for genuine patient engagement, and paying particular attention to communities with the most pressing medical needs and burden of disease. The task force additionally called for the creation of a Health Data Office and to develop a process for evaluating projects and transactions involving access to UC health data by outside parties. AI technologies rely upon data, and AI governance should be integrated with these ongoing UC data governance efforts.

Laws & Regulations Affecting AI in HR

California Assembly Bill 168 (AB 168)

---

[202] R. Gillon, "Medical ethics: four principles plus attention to scope", July 1994, BMJ, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2540719/pdf/bmj00449-0050.pdf.

[203] "ReCODE Health Tools," ReCODE, accessed Sept. 8, 2021, https://recode.health/tools/.

[204] Working Group survey of UC CIOs and CTOs, May 2021.

[205] Ibid.

[206] "Report to the President: President's Ad Hoc Task Force on Health Data Governance," January 26, 2018, https://www.ucop.edu/uc-health/reports-resources/health-data-governance-task-force-report.pdf.

AB 168 protects job candidates from being forced to reveal their prior salary to potential employers.[207] Specifically, this bill prohibits employers from requiring potential employees to reveal their prior salary to determine whether they will offer the applicant an offer or determine their salary. It also requires employers to provide the job pay scale, given reasonable requests. AB 168 does not prevent candidates from voluntarily revealing salary information, which can then be considered by employers if provided voluntarily. For UC, AB 168 may affect use of AI to determine salary.

California Assembly Concurrent Resolution 125

The California Assembly Concurrent Resolution 125 urges the Federal Government and California State Government to explore the use of emerging technologies, such as artificial intelligence, with the goal of expanding hiring pools, protecting against discrimination and bias in hiring, and ensuring that no candidates who are best suited for roles are denied opportunities on the basis of characteristics like race, gender, or socioeconomic status.[208] This bill bases this recommendation in the ability for emerging technologies to remedy the employment inequality that continues to permeate in our society, as well as the role of California as a leader in the pursuit of appropriate standards to deploy emerging technologies to reduce discrimination and expand opportunity.

Laws & Regulations Affecting AI in Policing

California Civil Codes

In California, there are civil codes that pertain to data collected through automatic license plate reader (ALPR) technology. Calif. Vehicle Code § 2413 puts in place requirements for ALPR data retention, use, and reporting for California Highway Patrol and Calif. Civil Codes § 1798.29 and 1798.90.51 define APLR data as personally identifiable information (PII) for breach notification purposes and establish requirements that entities using ALPR publish privacy policies related to data collection, use, and security processes. As such, UC data governance processes for ALPR data should follow established privacy and security safeguards and breach notification processes in place for PII.

Fourth Amendment

By deploying AI-enabled tools that collect data on individuals, applications may inadvertently affect individual's rights against unreasonable search and seizure protected under the Fourth Amendment.

AB 1215 - Moratorium on Facial Recognition

Biometric surveillance technologies have been shown to discriminate against people of color and women; thus, the moratorium was established to prevent further inaccuracies and mitigate the harms of pervasive surveillance.

---

[207] AB-168 Employers: Salary Information, 2017-2018 Reg. Sess. (Cal. 2017), https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180AB168.

[208] ACR-125 Bias and Discrimination in Hiring Reduction through New Technology, 2019-2020 Reg. Sess. (Cal. 2019), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200ACR125.

Computer Fraud and Abuse Act

Enacted in 1986, the Computer Fraud and Abuse Act (CFAA) prohibits accessing a computer without authorization or access in excess of authorization.[209] The CFAA is increasingly being considered in light of the use of web scraping to capture public data to train AI models. The legality of web scraping has come under scrutiny in a high-profile case involving Microsoft-owned LinkedIn and HiQ Labs, creator of the facial recognition system Clearview AI.[210] HiQ Labs scraped public profile data, including headshots, from LinkedIn to train its facial recognition system. Microsoft is claiming this violates the CFAA as it constitutes excess authorization to the platform and its data. For UC police, efforts to scrape data from public social media profiles may be in conflict with the CFAA.

Laws & Regulations Affecting AI in Student Experience

Student Privacy Laws - FERPA

In the United States, the primary statute governing privacy of student data is the Federal Education Rights and Privacy Act (FERPA).[211]FERPA protects the privacy of student "education records," which are broadly defined to include records that are directly related to a student and maintained by an educational institution such as the University or by a party acting for the institution, with certain specified exceptions.[212] As a practical matter, the large majority of records maintained about students at UC fall under FERPA protection.

FERPA sets out the baseline rule that personally identifiable information (PII) from a student's education record may not be disclosed, except with the student's written consent.[213] FERPA also sets forth a number of exceptions, instances in which the student's written consent is not required prior to disclosure of PII (e. g., UC is permitted to disclose PII from education records internally to school officials who have a "legitimate educational interest" in the data). Also, UC contractors can qualify as school officials if the contractors are performing work the university has outsourced to them, if other conditions are met.

Certain AI-enabled tools will not implicate PII of students because of the use of de-identified datasets for creating models. However, for any AI-enabled tool that makes use of student PII, the definition of a "legitimate educational interest" (LEI) will be important in applying FERPA protections. For example, the student experience section cites several AI-enabled tools that require use of student data to build effective predictive models, and then to actually apply the models to obtain actionable information (e.g., machine-aided admissions and financial aid decisions and automated interactions and/or interventions with respect to student success and mental health).

University campuses have discretion to reasonably determine a school official's LEI in PII from an education record. And campuses may have FERPA implementing procedures that further define (or

---

[209]"Computer Fraud and Abuse Act (CFAA)," NACDL, https://www.nacdl.org/Landing/ComputerFraudandAbuseAct.

[210] "Supreme Court Sends Web Scraping Case Back to Lower Court," June 14, 2021, Electronic Privacy Information Center, https://epic.org/2021/06/supreme-court-sends-web-scrapi.html

[211]20 U.S.C. § 1232g; see also 34 C.F.R. § 99 (2020) (providing guidance on the operation of the Family Education Rights and Privacy Act from the Department of Education).

[212] 34 C.F.R. § 99.3.

[213] 34 C.F.R. § 99.3.

narrow the scope of) LEI.[214] The LEI underlying many internal disclosures of student PII is obvious—there is no question that a teaching assistant must be allowed to provide personally identifiable grading information to the lead instructor for a course. That data transfer clearly furthers the student's education as evaluation of student coursework is a quintessential educational interest. The underlying LEI is less obvious when a staff member is using student PII to build AI models that will be used in the future for the benefit of other students. If the educational interest at stake pertains to future students only, personnel should carefully consider whether this meets the definition of LEI, and should consult policy experts and counsel.

Of course, in many instances, the first step in creating the AI model might be to strip the data of identifiers, such that it no longer constitutes PII and is then outside of FERPA's protections. However, personnel should consider whether this de-identification process must be done by the original data holder to avoid a disclosure of PII without LEI justification.

FERPA also includes exceptions to the baseline non-disclosure rule that may be applicable to AI-enabled tools.

- **Financial Aid**: FERPA permits disclosure of student PII for purposes of determining financial aid eligibility, amount, conditions, and enforcement.[215] Again, it is unclear whether this basis would cover disclosure for use with models to be applied in the future.
- **Developing, validating and administering predictive tests, administering student aid programs, and improving instruction**: FERPA permits disclosure of student PII to organizations conducting studies on behalf of educational institutions for these listed purposes, if certain conditions are met. One of these conditions calls for a written agreement and destruction of PII when the information is no longer needed for the purposes of the study, and calls for specifying the time period in which information must be destroyed. "Predictive tests" would traditionally refer to standardized testing or tests with similar predictive properties. It is unclear whether "predictive tests" would encompass a wide range of AI-enabled tools that have predictive intentions.
- **Directory Information**: FERPA permits disclosure of "directory information," unless a student has opted out from such disclosure. Each campus maintains its own list of directory information data points, within parameters established under systemwide policy.[216] Typical directory information data points include a student's name, home and email addresses, major, enrollment status and the like.

UC's Privacy Policies

In addition to compliance with FERPA and in connection with privacy of student data, UC must comply with its own internal policies relating to privacy. While UC has several policies that implicate

---

[214] "Disclosure of Information from Student Records," University of California, 2008, https://campuspol.berkeley.edu/policies/studentrecdisclosure.pdf.

[215] 34 C.F.R. § 99.31(a)(4)(i).

[216] "Policies applying to Campus Activities, Organizations and Students (PACAOS)," University of California, March 1, 2019, https://policy.ucop.edu/doc/2710533/PACAOS-130. Sec. 130.251.Complete list of directory information (which may be narrowed, but not expanded, by campuses, is: student's name, e-mail address, telephone numbers, date and place of birth, field(s) of study (including major, minor, concentration, specialization, and similar designations), dates of attendance, grade level, enrollment status (e.g., undergraduate or graduate, full time or part time), number of course units in which enrolled, degrees and honors received, the most recent previous educational institution attended, photo, participation in officially recognized activities, including intercollegiate athletics, and the name, weight, and height of participants on intercollegiate University athletic teams. https://policy.ucop.edu/doc/2710533/PACAOS-130

privacy, a key underpinning to UC's privacy programs is the 2010 Privacy and Information Security Initiative (PISI), which resulted in a UC Statement of Privacy Values and Privacy Principles stemming from the values.[217] PISI sets forth two fundamental aspects of privacy:

- **Autonomy Privacy**, which refers to "an individual's ability to conduct activities without concern of or actual observation" and
- **Information Privacy**, which refers to "the appropriate protection, use, and release of information about individuals."

Both realms of privacy affect deployment of AI-enabled tools. Some of the tools discussed in this report collect data on individuals, which is a form of observation of the individual (autonomy privacy). AI-enabled tools also implicate use of information about individuals (information privacy). Indeed, this report includes protection of privacy and security as one of the fundamental responsible principles for use of AI. The University's PISI helps to flesh out the details on what is meant by "privacy." Personnel involved in deploying AI at UC should be familiar with PISI and the privacy concepts it describes.

[217] "UC Statement of Privacy Values", University of California, accessed July 21, 2021, https://www.ucop.edu/ethics-compliance-audit-services/_files/compliance/uc-privacy-principles.pdf.