University of California Presidential Working Group on Artificial Intelligence Standing Council (AI Council)

Report of the Subcommittee on Transparency

June 28, 2024

Contents

Introduction

The AI Council's Subcommittee on Transparency (Subcommittee) is charged with developing approaches to promoting transparency to the University community and to the public on ways in which AI is being utilized or may be utilized within the University of California. Transparency in the use of AI enables the University to better evaluate potential risks and opportunities, study University experiences and outcomes, and to determine subsequent initiatives, such as the development of policy relating to responsible AI use that promotes efficiency, transparency, civil liberties, autonomy, and leads to equitable positive outcomes.

For FY2023–24, the Subcommittee sought to lay operational groundwork for articulating the University's uses of AI in a transparent manner. First, a short systemwide survey was used to quickly update the Council's understanding of patterns and trends around AI use, including generative AI, by the University since the 2021 Final Report. Second, a definition of AI uses that may yield a "high impact" was developed to extend the Final Report's identification and exploration of use cases that significantly affect an individual, organizational unit, and/or the University. We report on these activities and provide recommendations and observations to inform activity for the coming year.

Survey Work

In March 2024, the Subcommittee distributed a systemwide survey to University employees and faculty, intended to quickly identify changes to the landscape since the (pre-GenAI) 2021 report. The findings presented here are based on a convenience sampling resulting in 264 responses (Figure 1), with all UC locations participating.

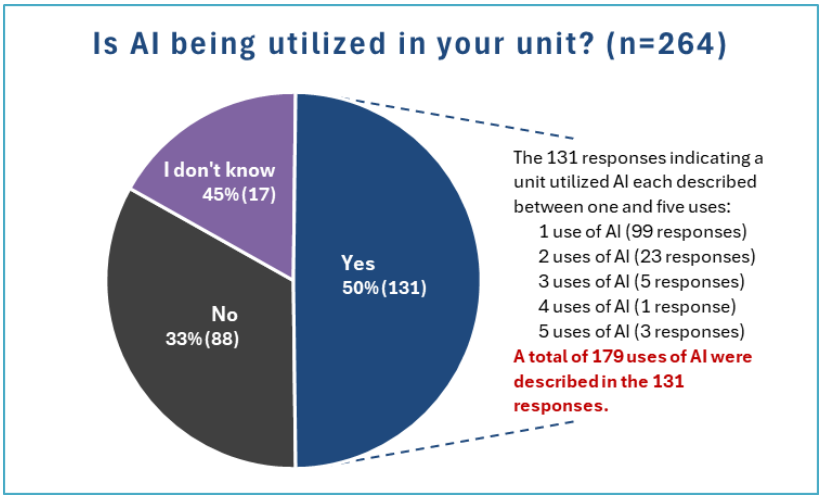

Figure 1. Summary of responses to "Is AI being utilized in your unit?"

Of the 264 responses received, half indicated their department or unit ("Unit") utilized AI, with a total of 205 uses of AI described. Though the majority of respondents described one or two uses, a small number of respondents provided three, four, and even five uses of AI in their Unit.

Respondents were asked to characterize each use of AI they described by any combination of "The AI provides information," "The AI makes a recommendation," and/or "The AI makes a decision" was most appropriate (Figure 2). Generally, if the AI made a recommendation, it also provided information; and if it made a decision, it also made a recommendation and provided information.

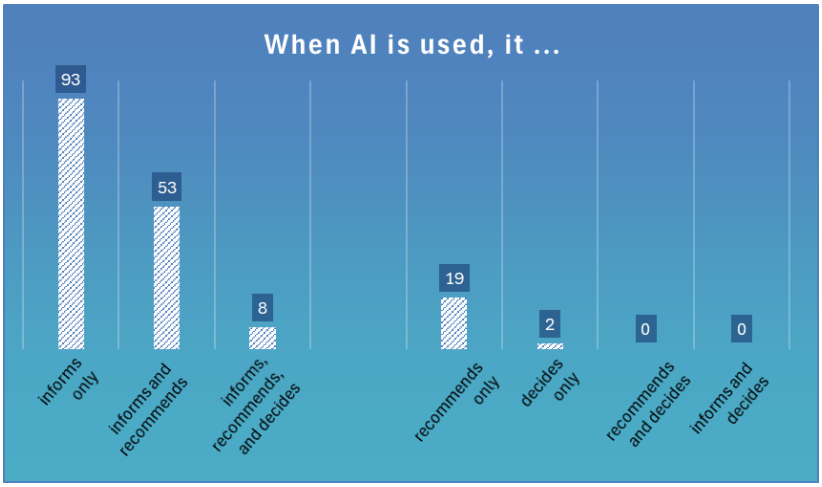

Figure 2. Whether AI is used to provide information, make a recommendation, and/or make a decision.

Fewer respondents indicated that their Unit supported AI for other units (Figure 3), but they described a total of 74 uses of AI they supported for others. Again, while the majority of respondents who supported an AI use supported only one AI use, there were some respondents that indicated that they supported more than one.

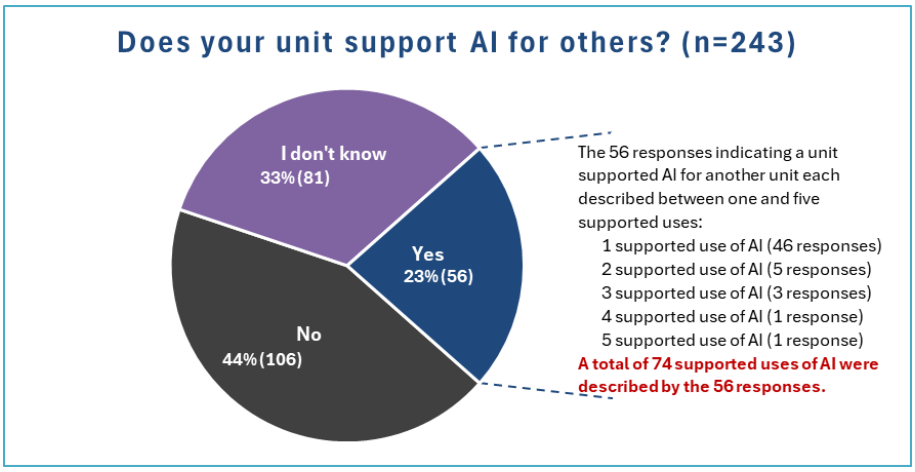

Figure 3. Summary of responses to "Does your unit support AI for others?"

Uses of AI by Domain

Survey respondents provided information on the use of AI across several domains and for various use cases. They also provided insights into the level of development of the use of AI, from experimentation or the launching of pilots, to more full-scale implementation. The chief domains are health, communications (internal and external), teaching, and research and analysis, as seen in Figure 4.

Domains	Modal Level of Development	Activities
Communications	Varies	Drafting, brainstorming, editing, creating targeted communications for different audiences, image editing and generation. Various chatbot support tools.
Health	Well-developed	Lots of activity, from patient scheduling, monitoring, and communications to record keeping, billing, note taking, and clinical assessments.
Teaching	Experimentation	Curricular and syllabus development, writing instruction, coding instruction, and training on how to use AI, along with instructional support for teachers and learners such as AI tutors or assistants
Research* **and Analysis**	Experimentation to implementation	Risk-modeling and other simulations; summarizing and transcribing meetings, summarizing professional, legal, or research literature; qualitative analysis of large amounts of textual data, business intelligence; drafting surveys

Figure 4. Uses of AI by domain and level of development.

Major Themes from Open-Ended Comments

Many comments were received in the open-ended comment fields of the survey. Four categories of comments emerged across all respondents: (1) clarification of use; (2) requests for policy guidance; (3) recommendations; and (4) general comments. Appendix B provides additional analysis of these comments.

Clarification of Use

Comments coded as "clarification," typically refer to respondents who included additional information on their Unit's AI utilization. For example, a respondent indicated that AI was being utilized by their unit to "provide information to students, grading." In the comments section, the respondent provided additional information to clarify this utilization by adding, "Making rubrics, grading assignments." Similarly, a comment coded as a clarification among respondents offering support to a unit also provided additional information regarding the type of support offered. For example, one respondent indicated that they provided support to a unit by "implementing a live survey…." In the comments section that followed, the respondent provided additional clarifying information regarding the survey instrument, specifically, the number of survey participants and categories used in data collection.

Though the survey did not specifically request comments relating to the development of guidance, many respondents expressed a need for such policy guidance. These respondents consistently characterized the need for guidance as "urgent" or as an "immediate need." Such comments focused on standards, ethics, privacy concerns, and vendor agreements/contracts. For example,

* The survey indicated interest only "in learning about administrative uses of AI and not research on AI." However, respondents were often provided information in free-form textboxes.

one respondent who utilizes AI in their Unit said there was a need to address ethical questions, such as "how AI can and should be used appropriately and ethically." Another requested guidance on privacy concerns: "strong guidance to students, faculty, and staff about the risks of providing personal or user data to public AI/LLM models." Respondents supporting AI uses in other units also requested guidance on privacy and ethics, as well as vendor agreements: "More guidance is needed around the UC stance on allowing downstream use of data…Data Use agreements have identified issues with the vendors…Issues with allowing vendor to de-identify the data."

Recommendations

Recommendations from respondents who utilize AI in their Units tend to suggest limiting or expanding AI utilization. For example, several recommendations focused on the need to include humans in decision-making, such as, "…we do not (yet) see a scenario where AI is robust enough…to make autonomous decisions – there needs to be a 'human in the loop' due to technical, clinical, and legal risks." Several recommendations identified a unit or group that they felt should be included in any future committee that might be focused on AI, such as librarians who are experts in AI, or the UC Davis Health Data Oversight Committee (HDOC).

Recommendations from respondents who support AI use by other units tended to suggest ways to limit (rather than expand) AI uses. To illustrate, one respondent who supports AI use in another unit indicated, "There needs to be human validation to check the outputs if [AI tools are being used] in situations that impact others." Another respondent suggested that while "generative AI can [summarize] complex information quickly…only a person who deeply knows the content of the many, complex clinical guidelines the tool is designed to summarize would be able to identify when the system generates a hallucination or meaningful omission."

General Comments

Most of the information provided by respondents in the open-ended comments box can be characterized as "general comments." For example, comments from respondents who utilize AI in their unit often included opinions, such as, "…there is a great deal of anxiety about the role LLMs will play moving forward…It is also true that students, faculty, and staff are largely transfixed by the moment, not using [AI systems/tools] out of that fear." Likewise, comments relating to respondents who indicated they offered support to units using AI, also included general opinions, such as, "AI is used for a variety of purposes, especially in research…[Tools] are evolving very rapidly and [it is] hard to keep up with all the new tech that is developed."

Limitations of Survey Results

Survey methodology can be found in Appendix B. However, we note several factors that likely influenced participation. First, concerns were expressed about how results would be used by the AI Council and UC generally, including to audit or monitor AI uses articulated in responses (particularly as the Council has members from legal and compliance functions). The possibility for disclosure of results under the California Public Records Act and the potential for misinterpretation by the public also likely impacted whether an individual responded to the survey at all, as well as their responses within the survey if they did respond. While it could not make guarantees, the Subcommittee was sensitive to these concerns, spending considerable effort to avoid asking for information that could readily identify a respondent and providing de-identified reporting if asked.

We also learned of other surveys about AI use at UC with specific perspectives, including surveys conducted by UC Health, UC Chief Information Officers, and research administration. UC Irvine was also in the midst of its own survey regarding AI use. While the Subcommittee tried to minimize

overlap with other efforts to the extent the timing of the Subcommittee's awareness permitted, survey confusion and fatigue likely had an impact on response to this survey.

Nonetheless, though the results can be considered preliminary — and no doubt already somewhat dated in the few months that have elapsed — they have strengthened the Subcommittee's confidence in understanding of patterns and trends of AI use across UC.

High Impact Work

As the Subcommittee intentionally crafted the survey as a convenience survey, intended to be answered in a relatively short period of time, the Subcommittee was also aware that the results for many of the uses would not provide significant information about uses that may implicate the UC Responsible Principles set forth in the UC Working Group AI Report of 2021. Thus, the Subcommittee decided that some survey results should be reviewed in further detail. Specifically, the Subcommittee sought to conduct its further review of AI use cases that could be of significant benefit or risk to the University, the people it serves, or University resources. Using current and then-pending legislation, regulations, and guidance as a guardrail (as of Fall 2023), including, but not limited to, the General Data Protection Regulation, the draft European Union Artificial Intelligence Act, and proposed regulations and legislation in California, the Subcommittee created a definition of "High Impact" use cases to review.

Definition of High Impact

AI is used for a specific sensitive purpose. While there is no exact list of use cases, sensitive purposes may include uses of AI relating to:

- Law enforcement (e.g., decisions to investigate or prosecute an alleged illegal activity)
- Responses to breaches or threats of information or physical security
- Employment
- Student admissions and financial aid
- Performance evaluations
- Access to social services of the University (e.g., education, housing, medical or mental health care, childcare, insurance)
- Clinical care
- Adjudication processes (e.g., student conduct, academic integrity)

An AI use will be "high impact" for purposes of this Working Group where the AI is used for a sensitive purpose and it recommends, or makes, a decision, that can significantly impact a person or University resources. A use can also be "high impact" where a person relies upon potentially inaccurate or biased output, taking an action that can have a significant impact on a person or University resource. A significant impact can include:

- Admission or denial of admission to UC;
- Enrollment or non-enrollment in a clinical trial, or provision or non-provision of clinical care that could lead to improved or negative health outcomes;
- Award or non-award of financial aid;
- Access on no access to student housing;
- Job offers or rejection of applications;
- Initiating or deciding not to initiate academic discipline proceedings;
- Responding to or not responding to physical or cybersecurity threats to University.

High Impact Uses

Using the above definition as a guide, and descriptions of AI uses respondents provided in survey results, the Subcommittee identified some of the uses that could yield significant benefit or risk to the University, the people it serves, or its resources. Subcommittee members then interviewed

stakeholders that could provide information about these uses. The methodology for identifying and interviewing stakeholders is summarized in Appendix C.

Based on the survey results and the definition of High Impact, the Subcommittee identified 57 entries that could be high impact uses of AI by UC. Many of these entries described similar uses. The entries initially flagged by the Subcommittee as high impact are described in Figure 5:

Type of Department Using Tool	General Summary
Health—Administrative	Forecasting number of beds occupied at hospitals; forecasting patients boarding in emergency departments; predicting likelihood of a patient missing an appointment. UC reliance on such tools could have an impact on patient care, access to social services, and employment decisions.
Health—Clinical/Patient-facing	Detection and risk tools; communications tools among providers, and between patients and providers; reconciling patient medications. UC reliance on such tools could have an impact on patient care.
Health—Compliance	Pharmacy diversion detection; monitoring access to electronic medical records; billing. UC reliance on such tools can impact how UC responds to cybersecurity threats and its compliance with legal obligations.
Campus—Academic	Outreach to learner groups for professional development courses; AI as a tool to improve learning by students; assisting in translation of textual content; identifying student success indicators; curriculum and research design; making course materials more accessible; grading tool. UC reliance on such tools can impact student academic performance.
Campus—Administrative	Forecasting enrollment; contract review and redlining; flagging potential instances of plagiarism. UC reliance on such tools can impact student admissions, performance, and evaluation.
Laboratories	Assessing traffic patterns and housing trends; evaluating trends in urban grids. UC reliance on such tools can impact not only access to social services for the University community, but for Californians at large.

Figure 5. Survey use cases initially flagged as high impact.

Due to time constraints, and difficulties in identifying stakeholders with knowledge of a specific use described in the Survey results, the Subcommittee was not able to learn more about each high impact use identified. Moreover, during interviews, some uses that appeared to meet the definition of "high impact" upon initial review, did not in fact, meet this definition when the Subcommittee was able to learn more about the use.

Some of the high impact uses identified in the Survey results are summarized in Appendix C.

Conclusions and Next Steps

Through the Survey, the Subcommittee was able to gain an understanding of the scope in which the University utilizes and would seek to utilize artificial intelligence. Many uses are potentially high impact, used for a purpose that could significantly impact a person or University resources. Survey results, as well as interviews with stakeholders also make clear that the University is seeking guidance on the factors that should be considered when the University seeks to employ AI, as well as when a third-party AI developer or provider seeks to use the University's data, whether pertaining to a University resource or a student, employee, patient, or research subject of the University. This call for guidance is even more pronounced when the AI use is high impact.

These results call for the Subcommittee to utilize the information it has gathered and focus on the following in 2024–2025:

1. A clarification of the benefits and risks to the use of AI, particularly in areas of high impact. This analysis will consider UC's risk appetite and assessment of risks.
2. The development of guidance, including guardrails, for promoting transparency in: (a) UC's utilization of AI, including the circumstances when UC should inform impacted and non-impacted individuals about its use; (b) UC's development of AI for non-research purposes, including the data used to develop such AI and how the AI is trained and addresses bias; (c) third party development and deployment of AI to UC, including how the AI is used and the data used to train the AI; and (d) third party use of UC data and resources for development and deployment of AI generally. This guidance will take into account:
    a. Promoting transparency for high impact uses of AI;
    b. The UC Responsible AI Principles set forth in the UC Presidential Working Group on AI Report; and
    c. The deliverables of other AI Council Subcommittees and UC AI work groups that may overlap with the development of this guidance, including vendor risk assessments and training.

Appendix A. Subcommittee Members

- **Hillary Noll Kalay**, Senior Principal Counsel, UC Legal, UCOP (Subcommittee Co-chair)
- **Kent Wada**, Chief Privacy Officer, UCLA (Subcommittee Co-chair)
- **Christine K. Cassel**, Senior Advisor for Strategy and Policy, Department of Medicine, UCSF
- **Coreen Harada**, Executive Director, Research & Innovation, UCOP
- **Mike Kennedy**, Deputy Chief Information Officer, UCR
- **Bill Maurer**, Dean, School of Social Sciences and Professor of Anthropology; Law; and Criminology, Law and Society, UCI
- **Camille Nebeker**, Professor in the Herbert Wertheim School of Public Health and Human Longevity Science, UCSD
- **Scott Seaborn**, Principal Investigator, Office of Ethics, Compliance and Audit Services, UCOP
- **Han Mi Yoon-Wu**, Associate Vice Provost and Executive Director of Undergraduate Admissions, Graduate, Undergraduate and Equity Affairs, UCOP
- **Zulema Valdez**, Interim Associate Vice Chancellor for Equity, Justice, and Inclusive Excellence and Professor of Sociology, UCM

Appendix B. Survey Details

Methodology

Before distributing the survey across University stakeholders, the Subcommittee piloted beta versions of the survey with members of the AI Council as well as with employees from Registrar and Undergraduate Admissions officers. Based upon initial results from this pilot, the Subcommittee revised the structure of the survey and the wording of survey questions.

The Subcommittee distributed the survey using a multi-prong approach. Members of the Subcommittee identified key stakeholders at the UC Office of the President (UCOP) level within Academic Affairs, External Relations and Communications, Ethics and Compliance and Audit Services, Office of Civil Rights, Finance, Procurement, Investments, Operations, UC Health, Academic Senate, Athletics, Legal, and the National Labs. UCOP contacts were asked to distribute to each of their campus stakeholders. The survey was also distributed to various systemwide listservs, reaching individuals that include procurement officers, privacy and compliance officers, contracts and grant officers, and others.

Individuals were provided with two weeks to respond to the survey. The Subcommittee received 264 responses to the survey with all locations participating. Among the 131 responses indicating that the respondent's Unit uses AI, the vast majority (88%) described one or two uses, with the remainder describing between three to five uses of AI.

Analysis of Comments

This analysis is focused on understanding respondents' open-ended responses to the following questions:

- For respondents who indicated that they engaged in an AI *use* within their unit, and provided an answer in the open-ended comments box, the survey question asked (for up to five AI uses within a unit), "The AI Council wants to learn more about how UC utilizes AI so that it can develop guidance for UC. If you have any other information about this utilization you think might be helpful for the AI Council, you can provide it here."
- For respondents who indicated that they provided AI *support* to another unit, and provided an answer in the open-ended comments box, the survey question asked (for up to five instances of AI support for other units), "The AI Council wants to learn more about how UC uses AI so that it can develop guidance for UC. If you have any other information about this use you think might be helpful for the AI Council, you can provide it here."

Responses included in the open-ended comments box were analyzed for general themes/topics. Four clear themes/topics emerged from responses (Figure 6), which comprised the four categories used for coding responses: (1) Clarification of Use; (2) Request for Policy Guidance; (3) Recommendations; and (4) General Comments.

Although there are some percentage differences between respondents who commented on AI use within their own Unit and those who offered AI support to other units, the four categories that emerged were the same for both groups. Most of the comments offered by respondents centered on providing general comments aimed at offering information that they believed would be helpful to the AI Council. Of all comments provided by respondents who indicated an AI use within their Unit, fully 44.4% offered a general comment, the largest response category. The second highest response category for this group was a clarification for a use identified in a prior question (33.3%).

Recommendations and Requests for Policy Guidance, which are comments aimed at voicing what respondents identify as a pressing need, were less common, at 15.9% and 6.3%, respectively. There were some differences between these respondents and those who provided AI support to other units. Respondents who provided support were equally likely to provide General Comments and Recommendations, at 33.3% each. One in five respondents who offered comments requested policy guidance (20.8%). Clarifications were least likely among this group, at 12.5%; this may be due to their "support" role, as they would not necessarily have additional clarifying information to provide.

Types of Responses	**Utilizes AI**	**Supports AI**
Clarification	21 (33.3%)	3 (12.5%)
Requests for Policy Guidance	4 (6.3%)	5 (20.8%)
Recommendations	10 (15.9%)	8 (33.3%)
General Comments	28 (44.4%)	8 (33.3%)
Total	63 (100%)	24 (100%)

Figure 6. Four Categories Observed in Open-Ended Comments Boxes

Survey Instrument

UNIVERSITY OF CALIFORNIA AI Council

Uses of Artificial Intelligence

The UC AI Council is asking for your help to better understand how Artificial Intelligence (AI) is used across the University of California. We would appreciate 10 minutes of your time to complete this short survey no later than March 22, 2024. You have been selected to answer this survey because of your role within UC, either as a person who has knowledge of how AI is used at UC, or as a person who can identify others within your Department or similar Departments at other UC Locations that may use or support the use of AI. As members of the AI Council, we acknowledge that we may not know who can provide valuable input. So, you are encouraged to further distribute to others within UC whom you believe may either use AI in their work for UC, or support others in their use of AI at UC.

The UC AI Council is composed of representatives from across the UC campuses and is charged with implementing recommendations from the UC Presidential Working Group on Artificial Intelligence Report issued in 2021. A goal of this survey is to initiate awareness of how AI is used across the greater UC community. This will enable the AI Council, acting for the benefit of the University, to better evaluate potential risks and opportunities, mitigate these risks and take advantage of these opportunities, and determine subsequent initiatives. The survey results will lead to an inventory of AI across the UC campuses and medical centers.

This survey prompts you to: (1) identify how your Unit *utilizes* AI and (2) how, if at all, you *support* other Units using AI, such as by negotiating or reviewing agreement terms, or supporting implementation. We are only interested in learning about administrative uses of AI and not research on AI.

Your answers will be reviewed by the UC AI Council. Unless we otherwise ask your Unit, any reports of uses of AI will only contain aggregated data. While the survey does not ask for your name, it does ask for your UC campus or medical center and your Department or Unit. We ask for this information so that we can identify duplicative uses and better understand AI uses across the University. Members of the AI Council may also reach out to your Unit to learn more about how your Unit uses AI.

If you have any questions about this survey or about the confidentiality of your answers, please contact us.

Sincerely,

Hillary Kalay, UCOP, Hillary.kalay@ucop.edu
Kent Wada, UCLA, kent@ucla.edu
Co-Chairs of the UC AI Council Subcommittee on Transparency

1. *What is your UC location?* *

   [LBNL | UCANR | UC Berkeley | UC Davis | UC Davis Health | UC Irvine | UC Irvine Health | UCLA | UCLA Health | UC Merced | UCOP | UC Riverside | UC Riverside Health | UC San Diego | UC San Diego Health | UC Santa Barbara | UC Santa Cruz | UCSF | UCSF Health]

2. *What is the name of your Department/Unit?* *

This Survey is divided into two sections. Section I asks how you utilize AI in your work. Section II asks if you support others in using AI (such as by negotiating AI vendor agreements, reviewing privacy, compliance, or legal terms, or helping implement AI systems for other Units).

Section I: Your *Utilization* of AI

This Section asks how, if at all, your Department/Unit *utilizes* AI in its work. This Section applies to you if your Unit utilizes an AI tool in performance of its duties.

Definition

AI refers to a tool or system that can perform tasks normally performed by a person:

- *AI can perform human-like tasks,* such as recognize images or speech, learn from data, identify patterns, generate written content or make decisions.
- *AI encompasses many kinds of technologies, such as machine learning (or: "ML"),* where algorithms learn through experience; and generative AI (or "gen AI, like ChatGPT), which generates new content or data based on a question or data given to the gen AI tool.

AI also includes using data collected from past and present events to predict the likelihood of specific outcomes.

3. *Is AI being utilized in your Unit?* *

   [Yes | No | I don't know.]

4. *Describe a purpose for which your Department/Unit utilizes AI in its work.* * *(You will have an opportunity to describe up to five.)*

   For example:

   - *AI canceling an appointment for a patient.*
   - *Provides a list of people to add to a recruitment.*
   - *Provides information about potential cybersecurity threats.*
   - *Flags potential academic dishonesty.*
   - *Provides information to students.*

5. *Does the AI provide information, make a recommendation and/or a decision? Please select all that apply.* *

   [The AI provides information. | The AI makes a recommendation. | The AI makes a decision.]

6. *The AI Council wants to learn more about how UC* utilizes *AI so that it can develop guidance for UC. If you have any other information about this* utilization *you think might be helpful for the AI Council, you can provide it here.*

7. *Do you have a second purpose for which your Department/Unit is currently utilizing AI?* *

   [Yes | No]

Section II: Your *Support* of Other Units Using of AI

This Section asks how, if at all, you support other Departments/Units at UC that utilize AI. This Section may apply to you if you negotiate contracts, implement new systems or products, or provide legal, compliance, security, risk, or privacy review of AI tools or systems.

Definition

AI refers to a tool or system that can perform tasks normally performed by a person:

- *AI can perform human-like tasks,* such as recognize images or speech, learn from data, identify patterns, generate written content or make decisions.
- *AI encompasses many kinds of technologies, such as machine learning (or: "ML"),* where algorithms learn through experience; and generative AI (or "gen AI, like ChatGPT), which generates new content or data based on a question or data given to the gen AI tool.

AI also includes using data collected from past and present events to predict the likelihood of specific outcomes.

8. *Do you support other Units that use AI?* *

   [Yes | No | I don’t know.]

9. *Describe an AI tool or system for which you have provided support to other Units. (You will have an opportunity to describe up to five.)* *

10. *If you know anything about how the tool is/was to be used, or the Department seeking to use it, please describe here.*

11. *The AI Council wants to learn more about how UC uses AI so that it can develop guidance for UC. If you have any other information about this use you think might be helpful for the AI Council, you can provide it here.*

12. *Do you have a second AI tool or system for which you provided support?* *

    [Yes | No | I don’t know.]

Appendix C. High Impact Interviews

Methodology

As the survey results identified locations and departments that responded to the survey, but not individuals by their name, Subcommittee members contacted individuals at the relevant departments to identify individuals who could provide more information about a given AI use. In some instances, the Subcommittee was not able to identify an individual with sufficient information about the AI use. In other instances, while the survey description of the AI use suggested that the AI use met the Subcommittee's definition of "High Impact," upon further review and discussion of the use, the Subcommittee determined that the use was not, in fact, high impact.

Subcommittee members sought to learn more about the AI use in the following areas during interviews:

A. AI Tool/System Generally
   - What it does
   - Creator (UC, industry)
   - Governed by a UC agreement?

B. Use of the AI Tool/System
   - Why UC uses/used
   - Considerations before UC adopted the tool
   - Decision-making process regarding use of the tool
   - Successes and drawbacks of the tool
   - Impact of tool on operations of Unit, decisions made, fiscal impact

C. Decision-making and Data
   - AI's role in any decision-making
   - Human oversight
   - Input of the AI tool
   - Data storage and security
   - Stakeholder awareness of tool

D. Training the AI Model
   - How was model trained?
   - UC data in training
   - Biases and mitigation of bias

E. Feedback regarding the AI use

Subcommittee members then summarized what they had learned about the AI uses and identified some of the UC Responsible Principles set forth in the AI Working Group Report that may be implicated by such AI uses:

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.

Interview Summaries

Interview 1: Medical Appointments

The 2021 UC AI Report highlighted UC San Diego Health's review of an algorithm to facilitate identification of potential no-shows to medical appointments. The Report described the tool as managing the possibility of no shows by "allowing for double-booking" of an appointment. The AI Council Transparency Subcommittee reviewed this use. The Subcommittee learned that in 2020, UCSD Health was looking for a solution to address high no-show rates to appointments. The electronic health record management system used by UCSD Health, and other UC locations offered the algorithm as one of the many offerings that can be "turned on" through its electronic health record system. The algorithm utilizes certain data as predictors to determine a prediction score for each patient, or their likelihood to show up for their appointment.

At the same time, UCSD Health was also looking into a reminder system, which would send a reminder to the patient's preferred method of communication and asks them to confirm their appointment. As implementation of the reminder system would take a year or two to implement, UCSD Health Enterprise AI Committee reviewed the algorithm, discussing its associated risks. Ultimately, the algorithm was presented to UCSD Health's operational leadership, where a plan was proposed on how to utilize the algorithm. UCSD Health agreed to introduce the algorithm on a six-month pilot basis. Clinic locations could view the prediction score for each patient, and based on that score, had the option of contacting the patient two to three days ahead of their appointment to determine whether they intended to come to their appointment. The tool was intended to serve as a means to remind patients about their appointments, not to be used a mechanism to allow for cancelling appointments or double booking. Within six months of use the algorithm, UCSD Health introduced the reminder system, which alleviates the need for offices to manually contact patients. As of February 2024, the algorithm is still "on," so offices still have access to the prediction score, but given the reminder system currently in place, it is unlikely offices use the score.

Principles Implicated

- **Appropriateness**: Intended use (i.e., sending patients a reminder) of the product takes into consideration the potential benefits and risks of overreliance upon the tool.
- **Fairness and Non-Discrimination**: The use of the product takes into consideration any potential bias the data alone may yield.

Interview 2: Population Health Model

The 2021 UC AI Report highlighted UCLA Health's 2018 development of a machine learning model to predict the risk of hospitalization and/or emergency department (ED) visits over the next 12 months in individual 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 management. 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.1 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 were 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 did not identify all at-risk patients, the team also provided a process whereby physicians have been able to utilize their own clinical judgment to identify additional high-risk patients. As of 2021, this model of care, termed the "Proactive Care Model," had been implemented in 50 UCLA Health primary care practices across Southern California and a preliminary review of the data showed a trend in (and potentially statistically significant) reduction in hospitalization and ED visits since implementation.

As of 2024, the model has been implemented across all primary care clinics at UCLA Health. The output of the model is the same. However, the number of identified patients assigned to the care management team has been scaled down to approximately 2,000 patients per quarter who are at the highest risk of hospitalization and ED visits. This helps the program provide greater focus and support to the highest risk patients. Once patients are identified by the model, the team reviews the health records of these patients to identify the appropriate support to offer, and contacts patients with the option to receive personalized proactive care.

Through a formal evaluation of the Proactive Care Model, the team also found that while the model did not result in significant change in ED visits, it did yield a 27 percent decrease in hospitalization. The team stresses that the model does not operate in a silo, and that it is one part of a larger clinical program, and does not replace clinical judgment, as a health care provider always has the option to refer a patient for intervention.

Since publication of the AI Report, the team also rebuilt the model for use across all campuses by using data from the UC Clinical Data Warehouse in collaboration with UC Health's Center for Data-driven Insights & Innovation.

UCLA Health has now used its new AI development capability and platform to build additional AI models. One new model predicts risk of rapid progression of kidney disease (to kidney failure or end-stage kidney disease) among patients with chronic kidney disease. It uses multiple data points in the EHR including lab results of patients such as glomerular filtration rate (GFR), quality of life metrics, and medical history. It allows early identification of patients and support to slow progression of kidney disease. Another model just completed complements the population risk model and the Proactive Care Program by predicting next year total cost of care for senior patients.

Principles Implicated

- **Appropriateness**: The use of the AI model is not the only way the patient care management team can conduct outreach to patients who may be at risk. Once patients are identified by the model, patients are contacted and asked if they would like to participate. The AI model does not automatically assign patients to reach additional outreach or care.
- **Fairness and Non-Discrimination**: The ability for physicians to self-select patients who could benefit from the patient care management team is one manner to address any potential bias that the data alone may yield.
- **Human Values**: Once identified by the model, patients are given the opportunity to choose to receive care from the patient care management team.

Interview 3: Contract Review

Since 2021, a UC San Diego contracts office has been using an AI tool to conduct its first line review of a narrow scope of redlined agreements from third parties. The Unit worked with a supplier to create a playbook, providing the supplier with their templates of this basic agreement type. This included redlines of this agreement type they had provided to third parties, rationales of why certain language may be unacceptable, and preferred clauses.

For several months, the Unit worked with the supplier, which would use their AI tool to redline an agreement; the Unit would review these AI-created redlines and provide feedback, allowing the tool to learn. After this pilot period, the Unit began using the tool. Once they received the first set of redlines from a third party, the team would send the agreement to the supplier to run the agreement through the AI tool. Once returned to the Unit, the contracts officer could review the redlines and rationales provided by the AI tool to review for accuracy and correct as needed. The goal of using this tool was to reduce the time of review for the first round of negotiation of these agreements. However, a few of the contract officers within the team did not trust the tool, spending almost as much time reviewing the redlines as if they had reviewed the redlines themselves. Despite this, they did find that it worked well for people new to the office, as one means of training junior officers about negotiation. They were also fully aware that the tool could not, and should not, be used as a crutch for individuals with less experience in negotiation. It also worked well for more seasoned officers, as they were able to quickly determine whether they agreed or disagreed with the redlines provided by the tool.

Principles Implicated

- **Appropriateness**: The potential benefit to contract review time must be weighed against the risk of overreliance on the tool. One way to consider the appropriateness of the tool is to consider the types of agreements in which AI should be used, focusing on less complex agreements that tend to have similar types of language and concerns.
- **Transparency**: Officers should have the ability to review the output of the AI tool and revise the redlines provided by the tool.
- **Accuracy and Reliability**: The AI tools should provide the same types of redlines even where the contract language may not match verbatim.

Interview 4: Workforce Access to Protected Health Information

Several UC health locations use artificial intelligence to surveil access to some of their electronic health records (EHR) systems to monitor for suspicious activity. The tool ingests detailed logs of user activity of an EHR system and runs the activity through its algorithm to assign a suspicion score to every access of an individual's EHR. Though the specific criteria and weight they are given by the algorithm is unknown, it likely considers factors such as whether others within the location have accessed the record; when such access has occurred; the duration of access to the record; when the patient has last been seen by the location; and how the record was searched. The suspicion score ranges numerically, with a low score being less "suspicious" than a higher score. The tool provides an assessment for each access event, providing information such as that described above. This information enables the compliance office to quickly gain context to prioritize their review of access events.

UC Health compliance departments have been using an AI tool for EHR privacy monitoring since 2019. Since then, locations routinely provide feedback to the tool, indicating after their human review of each event whether they determined the event to constitute a violation (indicating access that was likely inappropriate) or not, and whether the event, though flagged by the AI system as having a higher suspicion score, was actually a false positive. This information is used to refine and train each location's model.

The data from each specific UC location is only used to train the model for that location; the system does not combine data from locations, regardless of whether locations share the same instance of an electronic records system. The company also does not utilize UC Health data to train the models of other health systems.

For example, one UC Health location found that suspicion scores below a certain threshold have historically yielded "false positives," such that human review of the assessment and access event has historically not shown any activity that would raise suspicion of access to the EHR. By providing feedback to the company regarding what its human review yields of access events, the location has been able to train the system over time to flag more of the types of activity that may result in inappropriate access to an EHR.

Locations shared that it is the industry standard to use an AI model for EHR surveillance. Using AI for this purpose has eliminated a significant amount of manual work; it provides a more efficient way to review more information.

However, the tool is only one tool they use to monitor activity to EHRs, as the tool does not provide a 100 percent success rate, nor does it work in all EHR systems, nor well in all clinical and administrative uses of information. It is also not a substitute for human review of access events but intended as a complementary tool to identify and perform reviews of potential inappropriate access to the medical record.

Principles Implicated

- **Appropriateness**: The potential benefits to identifying inappropriate access to PHI through AI must be weighed against the possible risk of overreliance on the tool.
- **Accuracy and Reliability**: The underlying algorithm should be updated, with real-world observations and data provided to the tool, to ensure that it improves over time.

Interview 5: Scribing Technology

Following review by UC health data governance teams, at the beginning of 2024, several UC health locations began piloting the use of AI to transcribe patient-provider conversations and draft clinical visit notes for the provider to review, edit, and file to the electronic health record. This scribing technology is being piloted by a small number of providers at UC health locations. Ambient AI scribing is intended to address the issues presented by clinical notetaking during and after patient encounters: it aims to reduce the burden of documentation within and outside direct patient encounters and to improve provider-patient engagement during the patient's visit. UC health locations intend to review data collected during the pilot period, measuring its impact on a provider's time spent documenting, as well as its usage and accuracy across racial, ethnic, and primary language groups of providers and patients. Health care providers across the country have similarly begun to pilot and use this type of technology. See, e.g., Tierney, A. et al. "Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation." NEJM Catal. Innov. Care Deliv. 2024;5(3), available at https://catalyst.nejm.org/doi/full/10.1056/CAT.23.0404; Coiera, E. and Liu, S. "Evidence synthesis, digital scribes, and translational challenges for artificial intelligence in healthcare." Cell. Rep. Med. 2022 Dec 20; 3(12): 100860, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9798027/.

Principles Implicated

- **Appropriateness**: The potential benefits in reducing the burden of documentation and improving patient-provider engagement during clinical visits must be weighed against any potential for inaccuracies in transcription, as well as the privacy and security practices and protections of third-party AI scribe services.
- **Transparency**: Providers and patients should be informed when AI-enabled tools are being used, when transcribing conversations and when a provider is reviewing the AI draft note before it is finalized into the medical record. Locations should also consider transparency when filing the reviewed AI note to the medical record.
- **Accuracy, Reliability, and Safety**: The AI scribing technology should be measured for its accuracy.
- **Fairness and Non-Discrimination**: AI-enabled tools should be assessed for bias and discrimination, particularly with respect to its ability to transcribe conversations by non-native English speakers, as well as transcribing conversations where an interpreter or patient representative is present. Procedures should be put in place to proactively identify, mitigate, and remedy any potential inaccuracies that may result.
- **Privacy and Security**: The AI scribe services, where offered by third parties, must be provided in a manner that maximizes the privacy and security of all persons and personal data.
- **Human Values**: The AI scribing services, and the data ingested by the tools should be used by health care providers and any third parties in ways that support the agency of patients, providers, and others whose voices and data may be collected as part of the services. These tools and the data collected should be used in a manner that adheres to laws and principles of privacy.
- **Accountability**: UC and third-party scribing services should be held accountable for their development and use of AI systems in service provision in line with the above principles.

Interview 6: Pharmacy Diversion

Several UC Health locations utilize AI for pharmacy diversion detection for controlled substances. One AI model used for this purpose reviews activity, such as prescription trends by department/practice area and provider, how often a provider is administering a controlled substance compared to their peers, and patterns of retroactively updating patient charts. Based upon this information, the model assigns a suspicion score to each administration of a controlled substance, with higher scores signaling potentially more suspicious activity than lower scores. Once the AI system flags an event with a higher suspicion score, those events with higher suspicion scores are flagged for review by the location's pharmacy diversion team. Once reviewed, the team determines whether or not they need to escalate the event and take further action.

In addition, after reviewing events flagged by the AI model, the team then provides the tool with their assessment of the event, indicating whether the event constituted a diversion or not. This enables the tool to learn from each location, refining each location's model for its own use.

The diversion team expressed that the model works well in departments/practice areas with high utilization of controlled substances and larger patient populations. However, it does not work well in smaller departments, as a provider with a higher utilization of controlled substances (for example, due to their area of practice) in a smaller population of providers can drive up the average utilization for the department, making it more difficult to discern diversion from appropriate uses of controlled substances. It also does not work as well in identifying diversions in practice areas with extremely high uses of controlled substances, such as in anesthesia.

The UC Health location's diversion team noted that it has become industry standard to use AI for diversion detection. They shared that while the intent of using AI is to improve the time efficiency of their work, given the current gaps in using AI for diversion detection, they must still conduct significant manual review.

Principles Implicated

- **Appropriateness**: The potential benefits of detection diversion of controlled substances through AI must be weighed against the possible risk of overreliance and under reliance on the tool.
- **Accuracy and Reliability**: The underlying algorithm should be updated, with real-world observations and data provided to the tool, to ensure that it improves over time.
- **Privacy and Security**: The tool should be designed and used in a manner, which given the sensitivity of the data about employees, maximizes their privacy.

Interview 7: Reconciling Patient Medications

Several UC Health locations use AI to help reconcile outpatient medications when high risk patients are admitted to the hospital. UCSF has been using AI for this purpose this 2022.

Once a patient is admitted to the hospital or clinic, pharmacy technicians review each patient's medication history prior to admission. They review the patient's medical record, reviewing dispense data, refill history, and each medication's "sig," or the provider's directions to the patient to taking the medication. This enables them to identify the current medications of the patient and dosing. For example, if the records show that a patient was dispensed a 30-day supply of medication, but the pharmacy provided the patient with 60 pills, the technician can infer, based on information available to them, that the patient was prescribed to take 2 pills per day. The technician then completes the medication history and must confirm the accuracy of the medication history with two "best" sources, such as the patient, or pharmacy.

Rather than making these inferences on their own and manually transcribing that information into their records, the AI tool collects this data from a variety of medical records available for the patient, infers any missing information, and pools all such information to create an updated medication history. The tool clearly identifies when it has made such inferences, and only does so when the tool is certain that the inference is correct (as identified by the example above).

UCSF reports that the tool saves technician time and reduces the likelihood of transcription errors. However, they also note medications from smaller pharmacies may not be available for the tool to identify, and tool does not identify medications where the patient has paid in full rather than using insurance. However, as was the case prior to use of the AI tool, technicians must follow their standard practice and confirm any medication history with two best sources, such as the patient themselves, or the pharmacy.

Principles Implicated

- **Appropriateness**: The potential benefits to using AI here – reducing transcription errors and time of pharmacy technicians, and the risks of AI (potentially incorrect or missing information) are carefully evaluated, considering that the risks remain regardless of AI use.
- **Transparency**: Technicians are informed when AI-enabled tools are being used, as the tool identifies when it has made inferences based on the data available to it.
- **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. It is difficult to compare the accuracy of the tool to the true medication history of a patient, as patients may also underreport medication usage.

Interview 8: Enrollment Management

Two offices on a UC campus have collaborated to use machine learning (ML) to evaluate the probabilities for accepted offers and student success indicators based on aggregated probabilities. While the use of predictive analytics for enrollment is a standard practice, these collaborators determined that the accuracy of their original, logistic regression-based methodology had declined over time, and they needed to consider models with greater complexity and ensuring their methods minimized, if not eliminated, biases.

The goal of these collaborators was to achieve greater precision in achieving their campus enrollment targets. They noted achievement of their goal due to the availability of new and diverse data points for their models, alongside a responsive, near real-time process to monitor data for updates to the models based on changes in the environment.

The collaborators noted that keys to their success were (1) the importance of role clarity, in that only the data team handles data and conducts the analyses; and (2) collaboration between the team creating the model and those involved in the operational use/implementation of the analyses.

The team acknowledged areas for further study and discussion, particularly in relation to using this type of modeling and associated data in an ethical manner to prioritize or recommend academic skills development support for students.

Principles Implicated

1. **Accuracy, Reliability, and Safety:** Models are regularly assessed and updated as new and more robust data are available.
2. **Fairness and Non-Discrimination:** Models incorporate data from multiple sources with an aim to provide a holistic understanding of the student population.

Interview 9: Online Proctoring

A UC professional school has used a third-party remote proctoring tool that can:

- Lock a student's browser during exams to prevent access to unapproved websites.
    - The browser lock tool does not necessarily use artificial intelligence.
- Scan the test taker's room/environment using the students' webcams during remote exams to flag unapproved aides or others in the room that might assist the test taker (using artificial intelligence)
    - This use case was limited to approximately 3 percent of the overall use of the third-party remote proctoring tool and was higher during the pandemic. It has since been discontinued due to privacy concerns.
- Use facial recognition software to monitor the eye movements of test takers to flag the potential use of unapproved materials or unapproved assistance from a third party.

This use case also was limited to approximately 3 percent the overall use of the third-party remote proctoring tool and was higher during the pandemic. It has since been discontinued due to accessibility and privacy concerns.

Students taking courses using the third-party remote proctoring tool are provided advance notice that a remote proctoring tool will be used, in the syllabus, and are given the choice to opt out of its use. In cases where students opt out, they are provided with a live/zoom human proctoring option during exams. Use of the tool has decreased since students returned to in person classes from 1,000 exams during the pandemic to 100 this school year. The school administration is also considering discontinuing the use of the third-party remote proctoring tool due to its high costs: $7 per student per exam.

The professional school is considering additional use cases for artificial intelligence tools, including:

- A third-party word processing tool that uses artificial intelligence to force citations for all copy paste text and assists students in using permissible AI tools when granted permission from their instructors and institution, automatically creating citations, recording prompts, and sharing them with the instructors. A limited, pilot version of this use case will initiate in June 2024.
- A third-party Learning Tools Interoperability (LTI) plug-in which permits students to use AI on specific assignments, with the permission of the instructor, on a cost limited basis. The tool will use an application programming interface (API) to ensure that data provided by UC students is not used to train the third-party tool's public learning model and is maintained within a secure environment.
- A third-party AI tool that can be used to scan and flag student-teacher interactions to provide feedback to class instructors.
- A third-party AI tool that can create mind maps, generating up to hundreds of new ideas based on the input of a single idea.
- A third-party Learning Tools Interoperability (LTI) plug-in which will use AI as a virtual teaching assistant which can refer students to educational resources based on questions and inputs.

Principles Implicated

- **Transparency**: Students are informed that a remote proctoring service which uses artificial intelligence will be used in the syllabus and are given the opportunity to request an alternative proctoring method.
- **Fairness and Non-Discrimination**: The AI proctoring service should be assessed for bias and certain elements of the tool that have demonstrated bias, such as facial recognition software, should not be activated.
- **Privacy and Security**: Because the remote proctoring service uses AI to collect personal data, its use should be assessed by the University Privacy Office and a vendor security risk assessment should be conducted by the University Information Security Office.

Interview 10: Political Science Research Lab

A UC Political Science Research Lab is developing an AI-based process for reviewing and analyzing high volumes of text and image data sourced from external media companies. The objective of the project is to partner with UC Computer Science and Political Science researchers at various UC campuses to develop a Large Language Model (LLM) that can cull through and analyze thousands of text and image files sourced from external media sources. The analysis conducted by the LLM will be complex in that it will not just be asked to identify files or images but will also be asked to conduct follow up assessments on what it originally found. For example, when a political science researcher is evaluating the impact of a flood on a particular community, the researcher could ask the LLM to find images of the flood and then estimate the number of casualties based on what it found.

The Political Science Research Lab originally started this project with the objective of using a commercial generative AI tool such as ChatGPT. However, project leaders are now looking to utilize a homegrown (UC or other academic institution) tool, in order to:

- Ensure that research and analysis is conducted in a controlled environment.
- Provide an opportunity for interdisciplinary collaboration with other UC researchers (in this case, researchers across UC working on the intersection of Computer Vision and Political Science).
- Minimize costs: the use of commercial LLMs would be costly considering the complex analyses required for this project. Working with other academic institutions/researchers will allow for free (or minimal cost) utilization of the LLM and an exchange of data between researchers.

The Political Science Research Lab has already been in contact with three separate UC researchers who have developed LLMs/machine learning tools that could be used for this project. The model will be tested periodically against "human" review and analysis of the same set of text/images conducted by UC student researchers.

Principles Implicated

- **Shared Benefit and Prosperity**: This is a partnership with other UC academic researchers, and the sharing of image and text file data, as well as the analysis of this data by the LLM, will allow for much quicker and more effective analysis of data that can improve research outcomes and collaboration opportunities.
- **Human Values**: The analysis provided by LLM will be used to assist the Lab with its human rights initiatives, including the assessment of the human costs of political violence and the forcible displacement of impacted populations.
- **Accuracy, Safety and Reliability**: The LLM developed in partnership with the Political Science Research Lab and other UC researchers will be periodically tested against human analysis of the same data to ensure accuracy and reliability.

Interview 11: Student Health

A UC Student Health Center has used two separate third party AI tools in the clinical setting.

Clinical Patient Triage

The first use case involves Student Health Center nurses using the Elsevier ClinicalKey tool to assist with patient triage and to act as a "second opinion" safety net when initiating a treatment plan for a particular diagnosis or set of symptoms.

How it works: Student Health Center nursing staff input a patient's initial diagnosis and symptoms, ensuring that no identifying information is included [only a patient's age and symptoms are input], into ClinicalKey. ClinicalKey uses generative AI to review Elsevier's library of peer reviewed clinical papers and literature for articles relevant to the patient's particular symptoms, and based on its analysis of relevant articles, it provides the nursing staff with a suggested treatment plan. The response from ClinicalKey includes citations to the articles from which it sourced its recommendation. The nursing staff can also ask Clinical Key questions related to treatment plan, such as "Should the patient see a physician today or can this wait until next week?"

Human review of Clinical Key's responses: All interactions between Student Health Center staff and Clinical Key are reviewed daily by the Chief Medical Officer, before any its treatment recommendations are implemented. When necessary, the Chief Medical Officer will intervene or modify the recommendations.

Initial results: Though the Student Health Center has not conducted a comprehensive study of the efficacy of ClinicalKey in augmenting the existing clinical triage process, there is anecdotal evidence of Clinical Key suggesting an accurate diagnosis and treatment plan that had not been previously considered by nursing staff. For example, a patient presented with yellowing of the eyes and no other symptoms, and while this suggested a high bilirubin level and potential liver complications, one of the potential causes of this condition identified by ClinicalKey was mononucleosis. Testing revealed this diagnosis to be correct and a relevant treatment plan was initiated for the patient.

Transcription of clinician-patient discussions

The second use case involves the use of a third-party AI-based transcription tool to transcribe conversations between a patient and clinician during a treatment visit.

How it works: At the start of the visit, the clinician asks the patient if they consent to their conversation being transcribed by the third-party AI tool. If the patient consents, the service is turned on. As the clinician speaks, it transcribes the conversation. It does not conduct a "word for word" transcription. Rather, it uses AI to analyze the words and phrases spoken in real time and produces clinical descriptions of the day-to-day vernacular used by both parties. Clinicians using the tool have been trained not to say anything that could identify the patient or lead to the re-identification of the patient when combined with publicly available data.

Human review: Each clinician reviews the transcription with patient to ensure accuracy and makes necessary revisions before any data is entered into the patient's medical record.

Initial results: Anecdotal commentary from clinicians indicates that the use of the AI-based transcription tool has saved them time and improved the accuracy of the data entered into patient records.

Status note: This project was conducted on a pilot basis and is currently on hold pending additional review.

Principles Implicated

- **Appropriateness**: The Elsevier ClinicalKey tool is an excellent resource for assisting nurses in diagnosing and developing treatment plans for patients. However, it should not be used as the sole source for treatment recommendations. The Student Health Center ensures that a physician reviews each case and provides feedback regarding ClinicalKey recommendations to address this concern.
- **Accuracy and Reliability**: ClinicalKey recommendations should be assessed against human physician recommendations for the same patients to ensure ongoing efficacy and reliability.
- **Privacy and Security**: Patient identifiers are removed from inputs to ClinicalKey. Additionally, during the physician review process, a further screening for, and removal of, identifying data elements or those that could lead to re-identification, should occur. For the AI based transcription service, the Student Health Center will need to ensure that the third-party vendor providing the tool has been assessed by the university information security team and approved for use with HIPAA covered protected health information and P4 classified data.
- **Transparency**: The Student Health Center informs patients that it uses AI technologies for triage and transcription purposes and obtains their consent before using either the triage or transcription tool.

Interview 12: Alternative Image Text in Course Content

The Teaching and Learning Center at UC Santa Cruz utilizes the GPT-4 tool to generate alternative text descriptions (also known as "alt text") for images in course content, improving accessibility for students with and without disabilities. The team currently uses a paid ChatGPT account and plans to transition to ChatGPT Enterprise in the future once a UC-wide agreement is in place.

This optional service streamlines the process of creating alternative text, which was previously done manually, and only in instances when there was an approved academic accommodation from the Disability Resource Center. While some faculty members have raised concerns about intellectual property, the tool has been well-received for its efficiency and positive impact on accessibility.

The generated alternative text undergoes a two-step human review process: first by staff within the Teaching and Learning Center and then by the faculty member who requested the service. This ensures the accuracy and appropriateness of the generated content.

Principles Implicated

- **Appropriateness**: The use of GPT-4 for generating alternative text is considered appropriate and aligned with the goal of improving accessibility.
- **Accuracy, Safety, and Reliability**: The human review process ensures the accuracy and reliability of the generated alternative text.
- **Transparency**: The service is optional for faculty, and they are informed about the use of AI in generating alternative text.
- **Accountability**: The team at UC Santa Cruz is committed to transitioning to ChatGPT Enterprise in compliance with future UC-wide agreements.

Interview 13: Safety Training

Administrative & Residential Information Technology at UC Santa Barbara is leveraging AI tools to enhance the accessibility and effectiveness of safety training materials. Inspired by the work presented by UC Riverside the team utilizes DeepL, an AI-powered translation tool, to translate previously English-only safety training materials into Spanish. ElevenLabs, a text-to-speech AI platform, is then employed to generate natural-sounding voiceovers for the translated content.

To ensure accuracy and cultural sensitivity, a student employee fluent in Spanish provides feedback on both the translation and the selected AI voice. This human-in-the-loop approach has not only saved time for staff but also yielded more accurate translations and voiceovers compared to previous generation tools. Feedback from users has been overwhelmingly positive, highlighting the improved accessibility and effectiveness of the training materials.

This project not only expands the reach of essential safety training but also demonstrates a successful and ethical implementation of AI to foster inclusivity and equitable access to information across the UC system.

Principles Implicated

- **Shared Benefit and Prosperity**: By providing safety training in multiple languages, the UC system promotes a safer and more inclusive environment for all members of its diverse community.
- **Human Values**: The inclusion of a student employee in the review process ensures that the translated content and voiceovers are culturally appropriate and resonate with the intended audience.
- **Accuracy, Safety, and Reliability**: The team's dedication to human oversight of AI-generated content prioritizes the accuracy and effectiveness of critical safety information, resulting in a significant improvement over previous methods.