Planning a Center for Standards and Ethics in Artificial Intelligence

Pablo Rivas PABLO_RIVAS@BAYLOR.EDU
Department of Computer Science
Baylor University
Waco, TX 76798-97141, USA

Jorge Ortiz JORGE.ORTIZ@RUTGERS.EDU
Department of Electrical and Computer Engineering
Rutgers University
Piscataway, NJ 08854, USA

Daniel A. Diaz Pachon DDIAZ3@MIAMI.EDU
Division of Biostatistics
University of Miami
Coral Gables, FL 33124, USA

Laura N. Montoya LAURA@ACCEL.AI
Accel AI Institute
San Francisco, CA 94107, USA

Short Paper*

Abstract

Artificial intelligence (AI) is a transformative technology impacting many critical sectors of society. In many cases, AI-based technology has improved the quality of life of many communities. However, AI has had unwanted consequences in other cases, leading to a lack of trust and slow adoption of new AI technology. Many groups from academia, industry, and government have created AI ethics standards to protect consumers, regulate practices, and provide tools for responsible AI. These standards can help reestablish consumers' trust in AI and ease industry and government's adoption of AI, pushing forward innovation and profitable enterprises. We are now planning to establish a Center for Standards and Ethics in AI (CSEAI). The CSEAI will address the industry's need to navigate, adopt, and comply with the imminent sets of standards and new regulations for AI-based technology. The CSEAI aims to: 1) Provide its members with access to research in the areas of AI ethics and standards, 2) Provide training and information about current and upcoming regulations for AI-related technology, and 3) Train and mentor the next generation of professionals in AI standards and ethics. In this paper, we lay out our plans moving forward.

Keywords: AI Ethics, Standards for AI Ethics, Machine Learning

1. Overview

Various regulatory groups have recently produced only a few critical standards for artificial intelligence (AI) ethics Koene et al. (2018b,a); Bryson and Winfield (2017); however, the

*. This material is based upon work supported by the National Science Foundation under grant CNS-2136961. http://cseai.center

number of standards currently in production is unprecedented. Furthermore, the likelihood of such standards being adopted as lawful, recommended, or mandatory practice is very high Kerwer (2005). Any American industry producing any type of AI-based technology today will soon be forced to comply with these standards in order to protect the public and increase the trustworthiness of such products. The Center for Standards and Ethics in Artificial Intelligence (CSEAI) aims to provide industry the services necessary for the adoption of standards and ethical practices in AI through research, outreach, and education. CSEAI's mission is to collaborate with industry and government research partners to design AI protocols, procedures, and technologies that enable the design, implementation, and adoption of safe, effective, and ethical AI standards. The CSEAI will leverage the fact that all site directors are also minority leaders, providing a unique perspective in protecting underrepresented populations. Furthermore, the varied AI skillsets of the CSEAI site directors will enable the center to address a variety of fundamental research challenges associated with the responsible, equitable, traceable, reliable, and governable development of AI-fueled technologies. Furthermore, the CSEAI directors are leading research in the diverse areas of AI, including fair design, explainability of deep learning models, fair generative models for data augmentation, adversarial robustness testing, and self-supervised modeling Bui and Marks II (2021); Rivas (2020); Stenton and Rivas (2020); Guarino et al. (2020). Research in these topic areas will focus on the applicability and accessibility of AI in critical commercial and public industries such as healthcare, telemedicine, cybersecurity, defense, security, utilities, transportation, and more.

1.1 Center Vision and Mission

Vision: To improve the quality of human life, health, safety, and well-being of humanity by ensuring the safe, effective, and ethical incorporation of AI into society.
Mission: To further the field of AI by leveraging the research mindset of academia and the needs of industry, and to provide applicable, actionable, standard AI practices for the industry and academia that abide by the core tenets of fairness, accountability, and transparency while promoting AI research.

1.2 Key Unmet Industry Research Needs Driving Center Creation

AI development today proposes exciting advances and significant concerns. The widespread decision-making of AI introduces entirely new types of liability that industry and government experts have yet to fully characterize and are becoming of increasing concern to legal analysts. The inherent risks of self-driving vehicles are commonly understood Bojarski et al. (2016); however, the risks associated with AI use in lesser-discussed applications such as utilities have hindered their adoption in mission-critical applications, despite their clear and evident benefits. That lack of standardized practices limits AI's adoption in many industries. Companies are often required to evaluate and make their own judgment calls about introduction and use of AI. This has caused external standard working groups to be formed and address this issue Koene et al. (2018b,a); Eitel-Porter (2021); Siau and Wang (2020); Floridi et al. (2018). Efforts are underway to mitigate societal concerns about AI fairness, such as ISO/IEC JTC 1/SC 42 ANSI (2017), ANSI/CTA-2089.1 ANSI (2020), EO 13859 Trump (2019), or IEEE P7000-P7014 Koene et al. (2018b); Spiekermann (2017); Schiff

et al. (2020). Much of the standardization work is driven by high-value customers such as the US Department of Defense. The CSEAI directors are heavily involved in developing the IEEE P7000-series standards on AI ethics and, should the CSEAI be accepted as an NSF IUCRC, the director's skillset will have a more significant impact. Researchers at the CSEAI would work on standardized models for AI orthopraxy that promote and measure fairness and explainability in AI research and development Bellamy et al. (2018); Gunning (2017). Industry and government will need to develop, deploy, and use effective AI that meets requirements and norms by the standards community; the CSEAI will research and provide tools, protocols, and technologies that will satisfy such needs.

1.3 Economic Importance of Research Area

The McKinsey Global Institute reports that by 2030 AI could deliver an additional global economic output of $13 trillion per year Bughin et al. (2018). However, with the widespread knowledge and availability of AI comes a great risk of irresponsible use, making it a double-edged sword. Significant negative consequences, far beyond the loss of privacy or finances, can occur when AI is given decision-making power over human life in either medical or defense scenarios Shaw et al. (2020); Braun et al. (2020). In industries where there is the potential for loss of life or property, regulatory agencies have adopted strict guidelines for mitigating loss. For example, the Federal Aviation Administration (FAA) has issued five volumes of Federal Aviation Regulations (FARs) to promote safe aviation in the United States and is only one of the fifty titles comprising the US Code of Federal Regulations (CFR) Kraus (2008). The potential for loss of life or property in aviation is significant; however, that risk is effectively mitigated by the strict rules and guidelines placed by the FAA. This is evidenced by the >15 million flights per year in the US, resulting in only 10-15 accidents Srinivasan et al. (2019). The aviation industry contributes $1.6 trillion annually to the US economy, significantly less than what the AI industry is proposed to deliver Administration (2011). These things considered, the eventual regulation of AI in the private and public sectors seems imminent.

1.4 Center's Uniqueness from other IUCRCs and NSF-funded Centers

Existing NSF-funded centers are exploring deep learning and machine learning applications; however, our center will have a broader approach. For example, the Center for Advanced Electronics Through Machine Learning (CAEML) uses machine learning but does not focus on standardized practices; however, the CSEAI will focus on standards. The Center for Alternative Sustainable and Intelligent Computing (ASIC) focuses on architectures that will enable high-performance machine learning, but concerns about AI fairness are beyond its scope; however, fairness is of interest to the CSEAI. The Center for Big Learning (CBL) focuses on deep learning applications where AI explainability might be a secondary concern; however, the CSEAI is concerned with this. The CSEAI will develop applied AI ethics and explainable, standardized, and trustable AI practices. Some specific research areas that researchers at Baylor University (BU), Rutgers University (RU), and University of Miami (UM) can address include debiasing models with variational theory (BU); characterizing input distributions to identify and correct bias (BU, RU, UM); looking at model generalizability/robustness under data set shift (RU); explainability and regularization in the

context of the latter two (BU, RU); other cases of traceability, verification, and correction mechanisms, with a human in a loop or semi-automated (BU, RU, UM); and research of search problems (BU, UM).

2. Broader Impacts

The CSEAI will organize education and workforce development activities to grow the ethical AI/ML workforce, providing students and industry professionals opportunities. Our objectives are to: **(1)** *Broaden participation* in AI/ML standards and ethics careers through clearly-developed Standards and Ethics Proficiency Credentials (SEPCs) made available through virtual courses. **(2)** Create professionally-developed, widely-available AI ethics *educational materials* for undergraduate and graduate students to make a pathway towards a career in AI standards and ethics clearly and widely available. **(3)** Widely *advertise* information on ethics and standards in AI research, educational opportunities, and careers to a diverse audience to increase participation and broaden the diversity of those choosing an academic pathway toward an industry workforce career.

2.1 Undergraduate Education, Mentoring, and Opportunities

Education: To help undergraduate students be introduced to the standards and ethics in the AI field, the CSEAI investigators will create short video teaching modules (1-3 lessons) that infuse standards and ethics-related ML design into undergraduate courses. These modules will be offered for free on the CSEAI website as supplementary instruction that can be included in various undergraduate courses. Once past the planning phase, all CSEAI investigators teaching undergraduate courses will convene to plan to teach basic AI, ML, or data science curricula using standards and ethics in AI. Based on their expertise, investigators will be assigned one or more standards and ethics curriculum modules to develop. Each investigator will teach and record these modules so that they can be used easily in a flipped-classroom model Keengwe (2014), where asynchronous video lectures and practice problems are conducted as at-home work and active, group-based problem-solving activities are employed in the classroom Herreid and Schiller (2013); Bishop et al. (2013).
Mentoring: Along with classes, undergraduate students need to be mentored to choose a standards and ethics career pathway. CSEAI will operate as a coordinator to promote both internships and mentoring. We will form a workforce development team to coordinate with the center investigators to recruit a variety of students for summer Research Experience for Undergraduates (REU) opportunities at CSEAI sites Norton and Bahr (2004). Researchers and graduate-student research assistants will serve as direct mentors to the undergraduate students they recruit. It is known that students participating in REUs are more likely to pursue graduate school and advanced degrees Granger et al. (2006). The workforce development team will pursue opportunities for internships in standards and ethics jobs. A database will be made available to students at partner sites and on the CSEAI's website.
Internships: The CSEAI website will publish information about available government and industry internships in standards and ethics in AI from agencies such as NTIA, FCC, NASA, and DoD. Additionally, student academic internships available in standards and ethics research labs and REU sites will be advertised. To help students prepare to be competitive applicants, the CSEAI Academy will work with career experts from each Career Center

at the corresponding site to provide information helping students prepare a competitive application and be ready for interviews.

2.2 Graduate Education, Mentoring, and Opportunities

Shared Graduate Courses: After the planning phase, the CSEAI investigators will convene to develop two graduate course topics on CSEAI research. Each course will be co-taught by two or more investigators. These courses will be recorded as they are presented and made available online across CSEAI universities, and for other universities to use for a nominal fee. The goal of these courses is to encourage graduate students to pursue standards and ethics in AI research. Individual CSEAI directors will be responsible for initiating the process at their sites for having the courses formalized.

Standards and Ethics in AI Development Course Track: CSEAI will work with standards and ethics in AI development experts at the IEEE and ISO to determine a set of undergraduate and/or graduate courses that could best prepare individuals for a career in standards and ethics in AI. This content and description of courses will be provided on the CSEAI website, and CSEAI institutions will collaborate to develop course plans for standards and ethics proficiency. These courses could include current courses taught at CSEAI sites and any needed online courses designed by CSEAI faculty that cover the core tracks of the center to be offered at participating sites.

Graduate and Postdoctoral Associate Mentoring: All graduate students and postdoctoral research associates performing CSEAI research will have their daily activities managed by their supervisory researchers and participate in Standards and Ethics Proficiency Certification piloting, collaborative CSEAI activities, broadening participation, and mentoring undergraduates.

Collaborative Activities: To forestall the development of techno-social silos, collaborative center-wide learning will take place between CSEAI research labs, graduate students, and postdoctoral research associates will have cross-focal area mentors as well. Monthly virtual "lunch and learns" will focus on 1 to 2 focus areas Mawhinney (2010), with graduate students and postdoctoral research associates giving presentations on their research progress. Planned quarterly virtual "speed mentoring" activities will allow all early career researchers access and exposure to CSEAI investigators and Senior Personnel Cook et al. (2010).

Broadening Participation Activities: Graduate students and postdoctoral research associates will create short lessons and interactive experiments they can present to any public audience. They will videotape their presentation for editing and dissemination through CSEAI, the REU students network, partner educational institutions, and targeted social media.

Mentoring Undergraduates: Graduate students and postdoctoral scholars will participate in mentoring undergraduates in their research groups and while participating in the annual CSEAI Summer School, especially students from smaller colleges and underrepresented groups.

Figure 1: CSEAI summer school tracks.

2.3 CSEAI Summer School

The CSEAI also plans to have a summer workshop series based on CSEAI-related topics and industry needs, as shown in Figure 1. These will include applied AI ethics, AI ethics standards, compliance advice or best practices, hands-on tutorials, and an exclusive space for our CSEAI partners' training program. We plan to rotate rotating among our sites, beginning at the University of Miami. We will provide a diploma. The funds necessary for the event will be drawn from CSEAI memberships and attendee registrations. Student volunteers will help run this summer workshop, which will be a collaborative effort between the industry advisory board (IAB) and the Academic Leadership Team (ALT). We will leverage the investigators leadership and experience in organizing these kind of events Rivas (2019a, 2018, 2019b, 2021). Participants at the CSEAI summer school will: **(1)** Learn about the fundamentals of applied AI ethics, AI ethics standards, best practices, evaluation, and interpretation of standards. **(2)** See how ethical considerations and standards have been developed for a range of AI technology and applications. **(3)** Develop hands-on experience with machine learning techniques for addressing real-world, industry needs-based, applied ethical problems. **(4)** Investigate new best practices and emerging methods for standard compliance. **(5)** Network with industry workforce, researchers, students, and potential members.

3. Center Composition

The research areas that the center can potentially serve are shown in Figure 2. Every site plays an important role in serving industry with accessible information, training, research, and development. In the figure, investigators are listed by site and external collaborators and consultants are listed as [C]. All the faculty at each site are key to successfully execute the center's vision and mission.

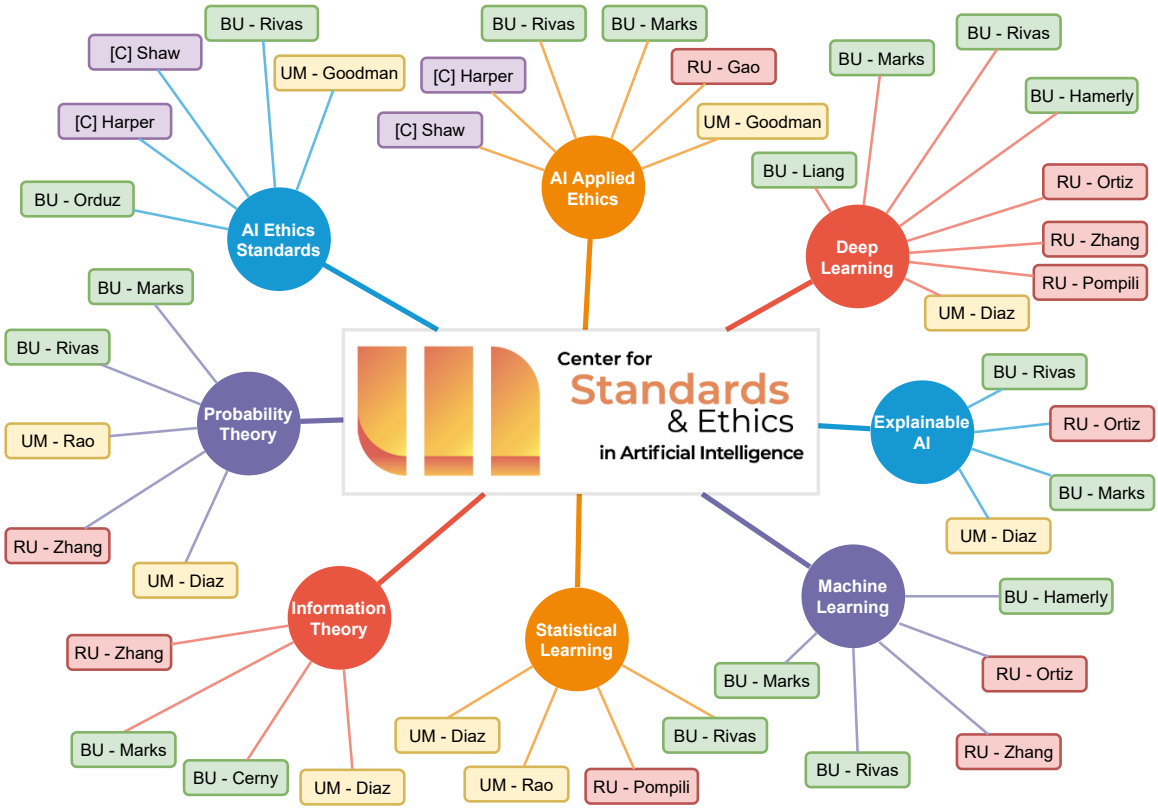

Figure 2: Research capabilities contribute to CSEAI's mission and vision.

4. Conclusions

Several regulatory groups have recently produced many critical standards for artificial intelligence (AI) ethics; however, the number of current production standards is unprecedented. Furthermore, the likelihood of such standards being adopted as lawful, recommended, or mandatory practice is very high. Any industry, particularly in the U.S., producing any type of AI-based technology today will soon have an obligation to comply with these standards to protect the public and increase trustworthiness in such products. The Center for Standards and Ethics in Artificial Intelligence (CSEAI) aims to provide industry the services necessary for adopting standards and ethical practices in AI through research, outreach, and education. CSEAI's mission is to collaborate with industry and government research partners to design AI protocols, procedures, and technologies that enable the design, implementation, and adoption of safe, effective, and ethical AI standards. The CSEAI will leverage the fact that all site directors are also minority leaders, providing a unique perspective in protecting underrepresented populations. Furthermore, the varied AI skillset of CSEAI site directors enables the center to address a variety of fundamental research challenges associated with the responsible, equitable, traceable, reliable, and governable development of AI-fueled technologies.

Acknowledgments

We would like to acknowledge support for this project from the National Science Foundation under grant CNS-2136961.

References

Federal Aviation Administration. The economic impact of civil aviation on the us economy. 2011.

ANSI. Iso/iec jtc 1/sc 42, 2017. URL `https://www.iso.org/committee/6794475.html`.

ANSI. Ansi/cta-2089.1-2020. *American National Standards Institute (ANSI)*, 2020.

Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilovic, et al. Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv preprint arXiv:1810.01943*, 2018.

Jacob Lowell Bishop, Matthew A Verleger, et al. The flipped classroom: A survey of the research. In *ASEE national conference proceedings, Atlanta, GA*, volume 30, pages 1–18, 2013.

Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

Matthias Braun, Patrik Hummel, Susanne Beck, and Peter Dabrock. Primer on an ethics of ai-based decision support systems in the clinic. *Journal of medical ethics*, 2020.

Joanna Bryson and Alan Winfield. Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50(5):116–119, 2017.

Jacques Bughin, Jeongmin Seong, James Manyika, Michael Chui, and Raoul Joshi. Notes from the ai frontier: Modeling the impact of ai on the world economy. *McKinsey Global Institute*, 4, 2018.

Justin Bui and Robert J Marks II. Symbiotic hybrid neural network watchdog for outlier detection. *arXiv preprint arXiv:2103.00582*, 2021.

David A Cook, Rebecca S Bahn, and Ronald Menaker. Speed mentoring: an innovative method to facilitate mentoring relationships. *Medical teacher*, 32(8):692–694, 2010.

Ray Eitel-Porter. Beyond the promise: implementing ethical ai. *AI and Ethics*, 1(1):73–80, 2021.

Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, et al. Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4):689–707, 2018.

Mary J Granger, Guy-Alain Amoussou, Miguel A Labrador, Sue Perry, and Kelly M Van Busum. Research experience for undergraduates: successes and challenges. *ACM SIGCSE Bulletin*, 38(1):558–559, 2006.

Michael Guarino, Pablo Rivas, and Casimer DeCusatis. Towards adversarially robust ddos-attack classification. In *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0285–0291. IEEE, 2020.

David Gunning. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*, 2(2), 2017.

Clyde Freeman Herreid and Nancy A Schiller. Case studies and the flipped classroom. *Journal of College Science Teaching*, 42(5):62–66, 2013.

Jared Keengwe. *Promoting active learning through the flipped classroom model*. IGI Global, 2014.

Dieter Kerwer. Rules that many use: standards and global regulation. *Governance*, 18(4): 611–632, 2005.

Ansgar Koene, Liz Dowthwaite, and Suchana Seth. Ieee p7003™ standard for algorithmic bias considerations: work in progress paper. In *Proceedings of the International Workshop on Software Fairness*, pages 38–41, 2018a.

Ansgar Koene, Adam Leon Smith, Takashi Egawa, Sukanya Mandalh, and Yohko Hatada. Ieee p70xx, establishing standards for ethical technology. *Proceedings of KDD, ExCeL London UK, August, 2018 (KDD'18)*, 2018b.

Theresa L Kraus. *The federal aviation administration: A historical perspective, 1903-2008*. US Department of Transportation, Federal Aviation Administration, 2008.

Lynnette Mawhinney. Let's lunch and learn: Professional knowledge sharing in teachers' lounges and other congregational spaces. *Teaching and Teacher Education*, 26(4):972–978, 2010.

M Grant Norton and David F Bahr. How to run a successful research experience for undergraduates (reu) site. *age*, 9:1, 2004.

Pablo Rivas. New york celebration for women in computing, 2018.

Pablo Rivas. Latinx in ai workshop at the international conference in machine learning (icml), 2019a.

Pablo Rivas. New york celebration for women in computing, 2019b.

Pablo Rivas. Ai orthopraxy: Towards a framework for ai that promotes fairness. In *2020 IEEE International Symposium on Technology and Society (ISTAS)*, pages 1–5, 2020.

Pablo Rivas. New york celebration for women in computing, 2021.

Daniel Schiff, Aladdin Ayesh, Laura Musikanski, and John C Havens. Ieee 7010: A new standard for assessing the well-being implications of artificial intelligence. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2746–2753. IEEE, 2020.

James A Shaw, Nayha Sethi, and Brian L Block. Five things every clinician should know about ai ethics in intensive care, 2020.

Keng Siau and Weiyu Wang. Artificial intelligence (ai) ethics: ethics of ai and ethical ai. *Journal of Database Management (JDM)*, 31(2):74–87, 2020.

Sarah Spiekermann. Ieee p7000—the first global standard process for addressing ethical concerns in system design. *Multidisciplinary Digital Publishing Institute Proceedings*, 1 (3):159, 2017.

Prabhakar Srinivasan, Venkataramana Nagarajan, and Sankaran Mahadevan. Mining and classifying aviation accident reports. In *AIAA Aviation 2019 Forum*, page 2938, 2019.

Eric Stenton and Pablo Rivas. Fine tuning a generative adversarial network's discriminator for student attrition prediction. In *22nd International Conference on Artificial Intelligence (ICAI 2019)*, page 13, 2020.

Donald Trump. Executive order 13859: Maintaining american leadership in artificial intelligence. *United States. Office of the Federal Register*, 2019.