

ACM SIGCAS

Computers and Society

Table of Contents

- 2 Who are we? and Membership Benefits
- 3 Introducing the SIGCAS Executive Committee
- 4 From the Editor - Richard Blumenthal
- 4 From the Chair: Welcome - Michelle Trim
- 5 We Need Your Short Pieces
- 6 Upcoming Events
- 6 Pursuing Social Good: An Overview of Short- and Long-Term Fairness in Classification
- Blossom Metevier
- 7 Nominations and Application Sought for Important SIGCAS Positions
- 8 GoodIT - Mikey Goldweber
- 8 Ethical Computing Instruction - Christina Sutcliffe
- 9 Talking About Visualising Sustainability to Enrich Computing: 1. Taking D'Amato to the Beach
- Samuel Mann & Ruth Myers
- 13 Third Annual SIGCAS Sowcase
- 14 Comparison of Rationales and Approaches of Artificial Intelligence Regulatory Policies in the U.S, EU and
China Draws Recommendations on a 'Fit for Purpose' Global AI Governance Mechanism
- Ao Kong, Weiyue Wu and Shaoshan Liu
- 16 Creative Coding as Method and Goal in a Gender Diverse First-Year Seminar - Paige Gulley
- 17 Project Unplug – A Personal Reflection on Rediscovering the Joy in Being Online - Michael James Heron
- 22 Fair Machine Learning Post Affirmative Action - Nick Perello & Przemyslaw Grabowicz
- 23 Classical Solutions for Quantum Challenges: An Introduction to Postquantum Cryptography - Brad Long
- 26 Opinion: Are You Really Competent To Be a Computer Scientist, a Computing Educator, or
Even to Use a Computer: The Missing Societal Competency - Richard Blumenthal
- 28 "All Watched Over by Machines," or AI Ethics, Surveillance, and Pluralism - Levi Pulford



**Association for
Computing Machinery**

Advancing Computing as a Science & Profession

Who Are We?

SIGCAS Computers and Society is the ACM Special Interest Group that addresses the social aspects and ethical consequences of widespread computer usage.

SIGCAS' main goals are to raise awareness about the impact that technology has on society, and to support and advance the efforts of those who are involved in this important work.

Our members are computer professionals from both industry and academia, as well as ethicists, psychologists, sociologists and others. We welcome students from a variety of disciplines. Our areas of involvement include computer ethics, universal access to computer technology, security, privacy, and reliability. We collaborate with other ACM bodies that are engaged in related work, such as COPE, USACM, SIGITE and SIGCSE.

ACM Computers & Society is an online publication accessible via the ACM Digital Library and the SIGCAS website. The newsletter aims to be an effective communication vehicle between the members of the group and to the outside world.

SIGCAS Computers and Society Readers and writers are invited to join and participate actively in this Special Interest Group.

Membership is open to all, for US\$25 per year, and to students for US\$10 per year. The link to join up can be found on our web site, at <http://www.sigcas.org>

Contribute. The editor invites contributions of all types of written material (such as articles, working papers, news, interviews, reports, book reviews, bibliographies of relevant literature and letters) on all aspects of computing that have a bearing on society and culture.

Please note that it is NOT a peer-reviewed publication. Submissions are checked for relevance, accessibility and basic suitability by the editors but not fully peer reviewed.

For the latest Call(s) for Papers, or instructions regarding formatting guidelines and copyright policy please see the website: <http://www.sigcas.org/>. Submissions may be sent to editors_sigcas@acm.org.

Copyright Notice By submitting your article or other material for distribution in this Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights:

- To publish in print on condition of acceptance by the editor.
- To digitize and post your article or other material in the electronic version of this publication.
- To include the article or other material in the ACM Digital Library and in any Digital Library related services.
- To allow users to make a personal copy of the article or other material for non-commercial, educational or research purposes.
- However, as a contributing author, you retain copyright to your article or other material and ACM will refer requests for republication directly to you.

Membership Benefits

- Subscription to the online publication ACM SIGCAS Computers and Society, which is published three to four times a year.
- Members have access to the full archive of the online publication and its printed predecessor in the ACM DL. Please see www.sigcas.org.
- Discounted registration fee for SIGCAS sponsored conferences and workshops. "In cooperation" sponsor of several ACM and non-ACM conferences related to SIGCAS' interests, including LIMITS.
- SIGCAS presents two awards each year: The Making a Difference Award and the SIGCAS Outstanding Service award.
- SIGCAS-ANNOUNCE mailing list: includes regular announcements of upcoming conferences and calls for participation. SIGCAS-Talk mailing list to enable member-member interactions and the committee will seek to stimulate discussion on this list amongst members. Subscription to the list is restricted to SIGCAS members and is optional for them.

ACM Computers and Society

Volume 52 • Number 2 • September, 2022 • www.sigcas.org

ACM Computers & Society (ISSN 0095-2737) is minimally published three times a year, normally in March, June, and September, with an optional issue in December, by ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701.

Richard Blumenthal Editor-in-Chief

Contributing Columnists
Richard Blumenthal
Douglas Schuler
Michelle Trim

Discord Moderators
John Bartucz
Benjamin Wade

SIGCAS Web Masters
Johanna Blumenthal
Alexandra Schultz

Write for Computers and Society?

The door is open, please walk through! We're all volunteers. Don't have time for a column, send your ideas, suggestions to: editors_sigcas@acm.org

Introducing The SIGCAS Executive Committee



Michelle Trim - SIGCAS Chair.

Dr. Michelle Trim is Senior Lecturer II and Informatics Program Director at the University of Massachusetts Amherst. In addition to designing and leading the Social Impacts of Computing course, she is currently PI and Co-PI on a number of grants that provide scholarships to students, increasing access to a computing education and providing programming that makes our context more welcoming to first generation college students and students from historically minoritized populations in computer science.



Samuel Mann - Vice Chair.

I strongly believe that computing is a powerful lever for positive change. I initially trained in geography and botany, and have dedicated my career to developing frameworks such that computing professionals can deliver on that promise of socioecological good. In the early 2000s I set out a framework for integrating sustainability into every course of study. Based on the transformation of education at Otago Polytechnic, it was later adopted for all polytechnic computing in New Zealand. Subsequent work has employed a “Transformation Mindset” and shows that even with the best of computing’s attempts for a sustainable future, we still have a lot to learn. I teach in a school of professional practice and work-based learning: Capable NZ Te Pūkenga, where we practice a heutagogical, self-determined approach to learning. Recent research is with indigenous groups, focused on understanding the potential and urgency of decolonising computing. I see great potential for SIGCAS to make a difference, but to do that we have to make ourselves relevant for all, especially the Global South.



Richard Blumenthal - Member-at-Large.

Greetings to the SIGCAS Community and thank you for your continued support. As Mikey notes, I also left a lucrative job in industry over two decades ago to focus on using my computing knowledge to more directly benefit society. My responsibilities with SIGCAS are currently in flux as I transition from the Editor-in-Chief of Computers and Society to our Member-at-Large. I am looking forward to serving the SIG in this new capacity. I am active in the Computing for the Social Good in Education (CSG-Ed) movement. For the past three years, I’ve served on the CS2023 Curricula Steering Committee, as the Chair of the Algorithmic Foundations subcommittee. I am also a member of the Executive Committee of the ACM Committee On Professional Ethics (COPE). Beyond ACM, I am a Professor and Chair of the Computer and Cyber Sciences Department at Regis University, in Denver Colorado. At Regis, I also contribute to our “Common Good Network”. I have a B.S., M.S., and Ph.D. in Computer Science from Lock Haven State, Rutgers University, and the University of Colorado, Boulder, respectively. I am very excited to be working with the new Board and look forward to helping make this the best SIG.



Mikey Goldweber - Past Chair.

Hello again to the SIGCAS community. As a past SIGCAS Chair, I hope many of you are aware of my passion and commitment to Computing for the Social Good. I left a high paying industry job in the mid-1980's (with the Porsche 911 to prove it) to seek a more personally rewarding career path. I landed in education after earning my PhD. However, I felt that being an educator was not quite enough; I needed to help my students see how computing can and should be used to improve society. Working in this area has taken me many places and afforded me the privilege of meeting many amazing colleagues doing amazing things. It also led me to SIGCAS, and after years of being a member, I stepped up into a leadership position. As the Past-Chair member of the Board I hope to continue offering my insights and time as our SIG moves forward to the challenges of the day.

From The Editor

By RICHARD BLUMENTHAL

Welcome to the second ACM SIGCAS *Computers and Society* issue of 2023. This issue is the first issue published under the new Executive Committee (see the Welcome below from our new Chair). As a result, I would like to call your attention to the Nominations for important SIGCAS positions on page 7 of this issue, which were also previously posted to SIGCAS announce and our Web site. Also posted to both "information channels" is the save-the-date announcement for our annual SIGCAS Showcase (see page 13 of this issue).

In addition to our recurring news announcements, this issues includes a summary of the recent GoodIT conference by Executive Committee member Mikey Goldweber.

In addition to welcoming Samuel Mann to the Executive Committee, his first installment, along with co-author Ruth Myers, of what is intended to be a long-running series of articles focused on trying to get computing to move beyond an understanding of sustainability that seems stuck in the 1980s. This was a fun article to read and I loved the dialogue format. As

an aside, my daughter Johanna, who has been active in SIGCAS and CSG-Ed, and I have been intending to write a series of articles for SIGCAS using a similar back-and-forth dialogue format to discuss computer and society issues. Hopefully, Sam and Ruth have inspired us to get started. I will admit, that I'm looking forward to rereading this sustainability article again since I'm sure I missed a lot on the first go-around.

Weiyue Wu and Shasoshan Liu, this time with Ao Kong, return as contributors with their article comparing approaches to AI regulatory policies among the U.S., E.U., and China. An easy to read, yet informative article.

Michel Heron, who has been a regular contributor to SIGCAS over the years, takes a completely different tack from his previous Scandal in Academia publications to reflect on his personal Project Unplug project. A very interesting self-ethnographic study, which may be a term I coined, but seem to remember reading.

A contribution by Brad Long follows that provides a nice overview of Quantum Computing and explores

the potential impact of this technology on society.

We also have a few Quick-Take and Short Piece submissions in this issue by Levi Pulford, Paige Gulley, Blossom Metevier, Christina Sutcliffe and

The issue concludes, as have all that I've edited, with my Parting Opinion column. As always, what I started to write evolved into something completely different. However, and as always, I take a for-all type position in order to try and spark some future conversation.

Finally, as a result of my recent election as a SIGCAS member-at-large and our by-laws specifying that no Committee member may hold two positions, I will be stepping down as the Editor-in-Chief of *Computers and Society*. It has been a pleasure serving our SIG in this capacity. As my "election campaign" promise, I look forward to finding another area of the SIG that I will endeavor to try and enhance. I look forward to continuing to work with you all and our new Executive Committee. — Rick

FROM THE CHAIR

Welcome

By MICHELLE TRIM

The SIGCAS membership elected a new executive committee this year, and our group is off to a strong start, working to develop activities to better understand the needs and wants of the membership. We have scheduled a retreat in January 2024 where we hope to develop a set of strategic goals and the beginnings of a shared vision for ways our organization can better serve our members, and our world, for the coming few years. As part of that work, we need more voices to join those collaborations, and we are hoping to increase our standing volunteer positions in the coming months. In particular, we are interested in recruiting volunteers to expand our editorial and social media communication efforts. If one of these appeals to you, please contact any member of the executive committee. Beyond improving our diversity and increasing our professional volunteer positions, we also want to expand opportunities to hear from our members about how SIGCAS can best serve you. The Town Hall, as part of our SIGCAS Showcase, is a great opportunity for sharing your ideas and concerns so that we can be in solidarity with each other. As we learn what SIGCAS can do for our members, we hope to also hear what our members would like to do for SIGCAS.

Why am I so concerned with growing our membership and getting our members more involved? The answer is that we are experiencing a new disruptive technology. I joined a panel at an event recently in New York City where internet founder, Vint Cerf asserted in his talk that what we need most today when it comes to the internet is "agency and accountability." His comment about agency struck a chord with me as I have been studying the intersection of technology and human agency since graduate school. What I took away from Cerf's definition of

'agency' in terms of the internet is that we must equip users and the public with the agency to protect themselves – to be informed in their consumption of content and services and to take meaningful action as a result. I thought of SIGCAS immediately as a place where some of the thinking about what it might look like to build user agency into designs of processes and tools. The rapid deployment and availability of AI-assisted technologies, ranging from decision-making systems to image and text generation, are ushering in a new constellation of environmental impacts and concerns. As our past chair stated in his parting opinion, "the intersection of computing and society contains the most important set of issues facing computer professionals today," and I have to agree. No other area of human activity has the same potential for a "no going back" kind of change at both the societal and the environmental level than that involving the intersection of unrestricted access to personal data and machine learning technologies.

I believe that one goal of an organization like SIGCAS must be to cultivate an ongoing conversation among a diverse community of professionals and students who care about computers and society. When we find ourselves among different kinds of people with similar concerns, with similar values, we discover several important things. One, we learn that we can learn from each other, regardless of the expertise we thought we had when we started. Two, we discover that sometimes what seems like the most immediate, the most urgent or the most serious way that computing and society intersect is not the same for all. In those moments we hear a heterogeneity of voices, and impacts that may reorder our prioritization of issues to be addressed. As a result, our identification of 'the problem' becomes both more accurate and more representative of a given concern's stakeholders.

Three, as we learn and complicate our sense of what the problems are, we find solidarity and a sense of shared empathy toward each other, and toward those in our communities more acutely impacted by particular technologies. This empathy leads to better designs, better systems, and more fair uses of computing that make our world better. In short, SIGCAS has the opportunity to be a collaborative space for shared values and collective action, raising awareness in colleagues, supervisors, students, researchers, and in industry.

As a new Chair of this organization, it is my sincere wish that we strive to grow our membership in diverse directions, making the texture of our communications layered, deep, compassionate, and intellectually rich. Now more than ever SIGCAS has the opportunity to facilitate the inclusion of underrepresented voices, to make space for people often marginalized in computing and in society. All of us can bring another chair to the table. Rather than try to speak for others different from us (whatever that means), every one of us can bring a chair, invite a colleague or a collaborator or a student to come to a SIGCAS event, and speak up for themselves. We all can do more to make sure all in our community feel welcome. I want to see our organization thrive and grow, as I believe SIGCAS should be a place where discussions are happening at the state of the art, focused on the changes to our world driven by rapidly expanding computing technology. Computing for good must begin with a recognition of who benefits and who loses. We can showcase those wonderful advances in computing that are equitably improving people's lives, and we can spotlight those computing systems and innovations that make life better for all.

We Need Your Short Pieces

Part of the Executive team's responsibility is to encourage more voices and varied perspectives on topics relevant to computers and society. Consequently, we at the "SIGCAS Publications Group" are seeking short pieces that are relevant, provocative, diverse, and unexpected for our issues of *Computers and Society*. We also hope they will be fun to write.

The possibilities, effects, implications, opportunities, challenges, myths, realities, and struggles related to computers and society that are being played out every day in millions of different ways are helping to determine who we are and where we are going. We want to capture at least some of that.

Your short piece could raise arguments, issues, critical questions, resource needs, current work, research, reviews, discussions, etc. etc. To that end we have developed a robust infrastructure of departments, divisions, bureaus, and other descriptive categories to help convey to you all that this is a vast, very formal and bureaucratized enterprise.

While some of the names may be fanciful we are optimistic that the articles they help characterize will be compelling, relevant, and influential.

We plan to experiment with this approach. We are currently planning to run several short pieces per issue. And we will probably add new departments at will. We also plan to be flexible but we do insist that these articles be short. (After all the SIGCAS Newsletter will still run longer pieces!) Shall we say 1,200 words max?

To be considered for the next newsletter please submit your short piece to the SIGCAS Newsletter Editor, Rick Blumenthal, editors_sigcas@acm.org, by December 15, 2023 (the subsequent issue will be published on December 31, 2023). Please include "Short Piece" in the subject line.

COMPUTERS AND SOCIETY AREAS OF INTEREST

- News From _____ (community, company, department, movement, country, sector, dimension, rain forest, or what-have-you)
- Your Resolution or Manifesto Goes Here Desk
- Not All is Wrong Department
- Systemic Racism & Black Lives Matter Studies and Reports
- Teaching about Computers and Society
- Social Responsibility in Computing Department
- Department of Development Studies
- Ominous Development Department
- What Could Possibly Go Wrong? Department
- Office of Emerging Technological Directions
- Voices of Practitioners and Younger Professionals
- Department of Diversity and Inclusion
- Climate, Biodiversity, and the Environment Department
- History Department (of SIGCAS and Computers and Society)
- Thrilling Adventures in Computing
- Looking at SIGCAS: Useful, Enlightening, Maddening or Other Influential Fiction, Poetry, Art, and Movies Related to Computing and Society Division
- Department of Technology Assessment
- War and Peace Studies Hall
- Help me work on myProject.dept
- Science Lab
- Religion and Spirituality Division
- Gender Notes
- Underscrutinized Implications Bureau
- Office of Expected and Unexpected Consequences
- What Should We Do Room
- Methods: How to do Computers and Society Group
- Annals of Agnotology
- Algorithms: Good, Bad, and Ugly
- What's a Professional Organization To Do Department
- Automating Evil: Office of Worst Practices
- Chronicles of Civic & Community Tech
- Department of Civic and Collective Intelligence
- Office of Technology Assessment
- Critics Corner (interviews, etc.)
- City Desk / Urban Studies
- Town and Country Consulate
- Point / Counterpoint Forum
- SIGCAS Agenda Development Department
- SIGCAS and Wicked Problems
- On the Job Department: SIGCAS and Employment
- SIGCAS and the Green New Deal
- Personal Perspective Department.
- You Can't Make This Up Department
- Design Perspectives and Perspectives on Design
- Student Voices Division
- Activism Sector
- Patterns of Computing Department
- Limits and Collapse Ministry
- Recent Reviews (books, articles, etc.)
- Steering Tech Department (policy and all the rest)
- Directions and Implications of Advanced Computing
- No Comment Department
- For or From The Archives
- Data and Datafication Office

Upcoming Events

The 13th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS 2024)

May 2-4, 2024 — Angers, France

A Conference that brings together researchers, designers, developers and practitioners interested in the advances and applications in the field of Smart City Information and Communication Technologies, Sustainability, Energy Aware Systems and Technologies (<https://smartgreens.scitevents.org/>).

ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS 2023)

August 16-19, 2023 — Cape Town, South Africa

Inspired by the broad agenda of the United Nations Sustainable Development Goals, COMPASS is an international forum for the presentation and publication of research in the domains of computing and sustainable societies, including fields such as artificial intelligence, data science, information systems, human-computer interaction, and engineering (<https://compass.acm.org/>).

Pursuing Social Good: An Overview of Short- and Long-term Fairness in Classification

BY BLOSSOM METEVIER

Machine learning (ML) models are increasingly being used to aid decision-making in high-risk applications. However, these models can perpetuate biases present in their training data or the systems in which they are integrated. When unaddressed, these biases can lead to harmful outcomes, such as misdiagnoses in healthcare [11], wrongful denials of loan applications [9], and over-policing of minority communities [2, 4]. Consequently, the fair ML community is dedicated to developing algorithms that minimize the influence of data and model bias.

Research in fair ML has predominantly focused on the classification setting, where predictive models, or classifiers, are tasked with categorizing data instances into predefined classes. For instance, in a lending application, a bank might use a classifier to determine the likelihood an applicant will repay a loan. Most fairness definitions for classification consider static fairness, i.e., the classifier's immediate impact on distinct groups or communities [3]. For example, a classifier adheres to the principle of statistical parity when the probability of (in)correctly predicting a specific outcome, such as loan repayment, is equivalent across demographic groups [6]. Other examples include fairness definitions that require equal performance across groups on standard evaluation metrics, such as equalized odds (equal true positive and false negative rates) [7], predictive equality (equal false negative rates) [5], and accuracy equality [1].

Static fairness methods serve as an important first step in the pursuit of long-term social good. Because they address immediate disparities, these fairness definitions often capture only a proxy measure for the impact models have on different communities. For instance, while a (static) fair classifier for loan decisions aims to improve a community's financial well-being, ensuring short-term fairness alone does not necessarily guarantee improvement [10]. Recent work, recognizing this limitation, goes beyond the myopic perspective of static fairness and directly focuses on long-term impact.

At a high level, strategies for mitigating long-term unfairness can be categorized into two groups:

methods that require an accurate model of the dynamics between classifier predictions and their subsequent effects on well-being (e.g., [8]), and (model-free) methods that assume no detailed knowledge of such dynamics, relying instead on observed data to reason about long-term well-being (e.g., [12]). Crucially, these methods support the idea from [10] that addressing long-term fairness requires insight into a classifier's broader societal impact.

In conclusion, while static fairness definitions provide an important snapshot of classifier bias, long-term fairness offers a more holistic lens. As predictive models become increasingly prevalent and influential in decision-making processes, it remains crucial for researchers and practitioners to rigorously address static and long-term fairness in their models.

References

- [1] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. 2017. Fairness in Criminal Justice Risk Assessments: The State of the Art.
- [2] Stephen Buranyi. 2017. Rise of the racist robots – how AI is learning all our worst impulses. *The Guardian*. (8 Aug. 2017) Retrieved 26 Oct. 2023 from <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>
- [3] Simon Caton and Christian Haas. 2020. Fairness in machine learning: A survey. *ACM Computing Surveys*.
- [4] Maurice Chammah. 2016. Policing the Future. *The Marshall Project*. (3 Feb. 2016) Retrieved 26 Oct. 2023 from <https://www.themarshallproject.org/2016/02/03/policing-the-future?ref=hp-2-111#.UyhBLnmlj>
- [5] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic Decision Making and the Cost of Fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*.
- [6] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness Through Awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12)*.
- [7] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS '16)*.
- [8] Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, et al. 2021. Towards long-term fairness in recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*, 445–453.
- [9] Kori Hale. 2021. A.I. Bias Caused 80% of Black Mortgage Applicants to be Denied. *Forbes*. (2 Sep. 2021) Retrieved 26 Oct. 2023 from <https://www.forbes.com/sites/korihale/2021/09/02/ai-bias-caused-80-of-black-mortgage-applicants-to-be-denied/?sh=3906206236fe>
- [10] Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. 2018. Delayed impact of fair machine learning. In *Proceedings of the International Conference on Machine Learning (ICML '18)*, 3150–3158. PMLR.
- [11] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew B.A. McDermott, Irene Y. Chen, and Marzyeh Ghassemi. 2021. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. In *Nature Medicine* 27, 12 (Dec. 2021), 2176–2182.
- [12] Aline Weber, Blossom Metevier, Yuriy Brun, Philip S. Thomas, and Bruno Castro da Silva. 2022. Enforcing delayed-impact fairness guarantees. *arXiv preprint arXiv:2208.11744*.

Blossom Metevier

UMass,

bmeterier@umass.edu

Nominations and Applications Sought for Important Computers and Society Positions

The SIGCAS Executive Committee seeks applications for multiple editorial positions for our *Computers and Society* publication.

Position: SIGCAS *Computers and Society* Editor-in-Chief

Deadline: November 15, 2023, with review of applications commencing October 16

SIGCAS brings together computer professionals, specialists in other fields, and the public-at-large to address concerns and raise awareness about the ethical and societal impact of computing technology. *Computers and Society** is an online publication indexed by the ACM Digital Library that publishes articles addressing current and historical social impacts of computing. The SIGCAS executive committee seeks applications for the role of SIGCAS *Computers and Society* Editor-in-Chief. This volunteer position is a member of the executive committee and oversees the solicitation of submissions, organizes the submission and review process, produces the newsletter proof, and ensures compliance with ACM formatting and style guidelines. The Editor-in-Chief position carries a 2 year, renewable, term limit. The Editor-in-Chief will produce 3-4 issues a year, and beginning in 2024, one of three issues will be exclusively hosting student work. Once this position is filled, the expectation is that *Computers and Society* will transition to 4 issues per year, with one student issue, one special topic issue, and 2 standard issues. The intention is to identify guest editors for the special and student issues on an annual, rotating basis.

Ideal candidates will have:

- A passion for facilitating a community of professionals in discourse on topics integral to understanding the intersection of computers and society
- >5 years working in a computing context, or post-tenure or well-established into teaching or industry positions
- High attention to detail
- Comfort collaborating with and communicating with individuals of diverse backgrounds
- An ability to discern the difference between weak writing that they agree with and well written/well-evidenced texts that they might not, as *Computers and Society* does not purport only one worldview
- Ingrained organizational skills that enable the balancing of competing deadlines and tasks efficiently
- Familiarity with using desktop publishing/layout software (such as Adobe InDesign (or its predecessor, Pagemaker, or similar application)
- Sensitivity to issues of accessibility as they relate to text and graphics

Opportunity: SIGCAS *Computers and Society* Guest Editor(s)

Deadline: November 15, with review of applications commencing after October 16.

The SIGCAS executive committee seeks applications for the role of Guest Editor for the Student Issue and Guest Editor for the Special Topics issue. The Guest Editor solicits, edits, and creates the proof for the content of a single issue, including providing a foreword or editor's note that thematically characterizes the submissions appearing in that issue. The Guest Editor reports to the *Computers and Society* Editor-in-Chief who will be responsible for the final formatting, processing, and publication of the issue. These are rolling positions, with roughly one of each type (special and student) releasing per year. Applicants for the Guest Editor for Special Topics position should submit with their application materials a 400 word abstract describing the theme they intend to pursue for their issue.

Ideal candidates will have:

- A passion for an aspect of the intersection of computers and society and engaging in its related professional or academic discourse
- >2 years working or studying in a post-undergraduate computing context. PhD students are eligible and invited to apply.
- High attention to detail
- Ability to energetically solicit submissions from diverse perspectives
- Experience communicating with individuals of diverse backgrounds
- An ability to engage the interest and productivity of potential authors
- Ingrained organizational skills that enable the efficient balancing of competing deadlines and tasks

Interested

To express interest in any of these volunteer positions, please send a cover letter and C.V. to michelletrim@umass.edu. All applications will be reviewed by the SIGCAS executive committee. We hope to fill these positions soon, so please don't wait to let us know that you might be interested!

*Please note that *Computers and Society* is not currently a peer-reviewed publication.)

GoodIT

By MIKEY GOLDWEBER

The Third ACM International Conference on Information Technology for Social Good (GoodIT) was held September 6-8, 2023, in Lisbon, Portugal. GoodIT is one of the primary conferences sponsored by ACM SIGCAS. GoodIT began in 2021 by a small cadre of mostly Italian-based academics. In just two years, GoodIT has attracted paper submissions and attendees from not only across Europe, but Africa, North America, and Asia. This year's conference was attended by approximately 100 delegates.

While the first and second iterations of GoodIT were hybrid, this year's conference was strictly an in-person affair. The conference was also run using a workshop style format: only one track, all presentations were plenary. However, given the growth of this community, future years may very well expand to a multiple track format.

In the words of the conference's chairs: GoodIT focuses on the application of information technologies (IT) to social good. Social good is typically defined as something that provides a benefit to the general public. In this context, clean air, clean water, Internet connection, education, and healthcare are all good examples of social goods. However, new media innovations and the explosion of online communities have added new meaning to the term. Social good is now about global citizens uniting to unlock the potential of individuals, technology, and collaboration to create a positive societal impact.

The conference attracted 144 submissions: 119 full papers (both regular and special tracks), 20 works in progress, and 5 PhD submissions (works in progress from PhD students). From this set, 54 full papers, 9 works in progress, and 4 PhD submissions were accepted via a single blind reviewing process with three reviews per submission. Hence, the acceptance rate for full papers was 45%.

As with the previous instances of GoodIT, attendees came away astounded by the variety and creativity of the projects our colleagues are working on. To get a sense of the range of topics/projects, consider the following subset:

- A tool to visualize the impact of the war in The Ukraine on climate change.
 - A gamification tool to provide auto drivers with feedback to encourage better sustainability practices.
 - An elderly-person fall detection system using ultra-wide bandwidth radars.
 - A facial recognition project to detect abuse of foreign laborers in the Taiwanese fishing fleet.
 - Two different approaches to detecting and addressing the threat from malaria carrying mosquitoes. A related paper won the best-paper award at GoodIT 2021.
 - Leveraging SmartMirror technology to improve home energy efficiency.
 - A project exploring a human-robot collaboration in the formulation of personalized medicine.
- There were also a variety of special tracks. These included:
- AI on networks for social good
 - Green IoT & environmental monitoring systems
 - Blockchain and decentralized technologies for social good
 - Citizen-Centric AI
 - Prosocial interactive digital narratives

- Games as entertainment and education instruments

- Networking and applications for mobile objects in land, water, and sky

It was hard to tease out any predominant themes from the conference. Issues related to sustainability were prominent, as was the application of machine learning to a many different domains. Other repeated themes included smart(er) cities, applications of blockchain technologies to benefit traditionally disenfranchised communities, and applications targeted to aging populations. This final topic was of particular interest to me, being what in the USA is labeled an Eisenhower baby (born prior to 1961).

Speaking of age, not only has the size of GoodIT grown, but the community looks and feels younger as well. Hopefully, this is a good sign regarding the long-term health of this community (and the planet).

The opening keynote was delivered by Michela Magas, the 2017 recipient of the European Woman Innovator of the Year. Ms. Magas is known for bridging design and technology, academic research and industry. Her talk titled "Tech for radical inclusion: creating spaces for common understanding" illustrated the advantages when one mixes together communities that do not normally interact.

The closing keynote was delivered by Daniele Quercia, the Director of Responsible AI at Nokia Bell Labs Cambridge (UK) and Professor of Urban Informatics at the Center for Urban Science and Progress at King's College London. Dr. Quercia's talk was titled "Insider Stories: Analyzing Stress, Depression, and Staff Welfare at Major US Companies from Online Reviews."

This talk presented results from his project of mining 440K company reviews published during twelve successive years on GlassDoor, and developing state-of-the-art deep-learning frameworks to accurately extract mentions of: stress and internal sustainability efforts.

Finally, it should be observed that all the attendees enjoyed the glorious hospitality (and excellent weather) of Lisbon in general and our local hosts in particular. The food highlighted the local cuisine, and yes, the port was plentiful at the receptions. Probably the only black mark for the whole event was the substandard coffee (by European standards) that was mistakenly provided. I, along with all the delegates I spoke with are already looking forward to GoodIT 2024 which is scheduled to take place Sept 4-6, 2024 in Bremen, Germany.

Mikey Goldweber

Dennison University,
Granville, Ohio, USA

samuel.mann, ruth.myers@op.ac.nz



Ethical Computing Instruction

By CHRISTINA SUTCLIFFE

Keywords: ethical computing, teaching computing, computing undergraduate students

"Why weren't we taught this earlier?" This question, posed by a college Junior in response to discussions of Ethical Computing, gave me hope. After all, it demonstrates the underlying frustration the student is experiencing while trying to internalize ethical concepts and align them with previous instruction. I wish all my computer science students faced that dilemma.

Instead, as I teach my way through CISC 305 – Social Issues in Computing, I get cynical observations on "the way it is" from about half my students – young adults with only two decades experience behind them. Despite the myriad of articles in the news about unethical computing [7], some students express that these instances are the exception and won't affect them or their career. And that raises the question: Why is computing instruction making these students evasive about ethical consideration?

Perhaps this dismissive attitude can be attributed to the age of the students? I don't think so. You see, I taught a sister course to CISC 305 last year – to nursing students (NURS 317). At UMass Amherst all students are required [3] to take a "Writing in the Discipline" course their junior year that teaches them to "develop critical awareness about writing within discipline-specific academic and/or professional communities." In both cases, these courses incorporated the ethical ramifications that could be faced within these professions.

In NURS 317, though, the ethical dilemmas we read and discussed in class were, without exception, already engrained in minds of the prospective nurses. Students of the same age range as those in CISC 305 were prepared to deal with topics ranging from abortion to euthanasia.

Okay, then maybe the attitude has to do with the practical application of their learning? But that can't be it either – as AI is being programmed to determine which patients receive a kidney transplant [6]; programmed robots are assisting doctors in surgery [1]; and AI chatbots are diagnosing medical conditions [4] in lieu of professionals.

Personally, I think the answer lies in the concept of the theoretical. Computing exists, after all, on a computer screen – and, as such, there may be little-to-no interaction between the developer and the end-user. So, the code can be theoretically divorced from its application – and any "failures" can be statistically nullified below a certain percentage of successes. This is certainly the case with one big-name company [2], which states on their website that budding developers can "use your development skills to take on global sustainability issues, win prizes, and have your solution deployed" – and yet it fails to include an specific ethical directive despite the fact that the company's most recent medical AI tool has been the subject of multiple studies [5] that point towards possibly inaccurate diagnosis across age, gender and socioeconomic status.

— continued on page 13 —

Talking About Visualising Sustainability to Enrich Computing: 1. Taking D'Amato to the Beach

By SAMUEL MANN & RUTH MYERS

Keywords: sustainability, diagram, representation, creative.
Categories: Social and professional topics → Sustainability.

In this series we take a significant contribution to the visual representation of sustainability, and probe it for insights for computing. In a post-structuralist approach, we present this as a dialogue.

Sam: Let's introduce ourselves first. I strongly believe that computing has the power to bring about positive change. My background in geography, computing and botany has enabled me to dedicate my career to developing frameworks to allow computing professionals to deliver on this promise of socioecological good.

Ruth: Socioecological, what's that?

Sam: We're so familiar with the term socio-economic to describe goals and processes related to people and business, but that externalises the environment, I like the reframing inherent in strong sustainability that the economy is a servant of the socioecological system. The more usual term is social-ecological but I like to reinforce the holistic nature, and it's easier to say. Your turn.

Ruth: I'm interested in creative practice as a means to open up spaces for thinking things differently. Artists are compelled to work within the messy complexities of contexts, with little guide ropes or resources, and this helps us pay attention. I see creative practice as available and helpful when thinking about change. In my current practice I am trying to listen and learn from place. Why are we here?

Sam: I think that computing's engagement with sustainability has been rather narrow. We've fallen into a transactional approach that focusses almost entirely on energy efficiency or we call it environmental sustainability – as if that were a sensible notion, the whole premise of sustainability is the inseparability of systems – but in doing so we miss the depth and diversity of sustainability science. If we can deepen computing's engagement with richer understandings of sustainability, then computing can have a much more important positive impact. And I think exploring diagrams is a good vehicle for that conversation.

Ruth: As an artist these diagrams are new to me. Why the diagrams?

Sam: Several years ago, I was working with two communication students on a teaching resource that would double as promotion material for our institution. The polytechnic had adopted a mission of "Every graduate may think and act as a sustainable practitioner". We had the idea of superimposing models of sustainable practice over the glossy marketing images – as a "sustainable lens" showing that every image, every event, every practice, was inherently about sustainability. I left the designers to it, with the instruction that they should adapt published diagrams of sustainability for the overprint images, but not to use a pillar model or Venn diagram. After a few weeks they phoned to say they couldn't find any other diagrams, so I started making a collection for them. After I got to 150 different diagrams of sustainability I wrote a book, but I couldn't stop, and the collection now has 800 images.

Ruth: It's a lot. I have to admit it took me a while to get inside the diagrams, navigate their abstract

thinking, but once I did, I started to see how they offered a way for collective thinking. How they could point out all the 'leftouts' of our thinking, our approaches to learning, our unquestioning of contexts and lenses, and especially of process and relationships. I like the questions we start asking ourselves.

I found myself wanting to 'world' the diagrams, bring a sensual, embodied understanding like Brady's 'artful-science' which accommodates the "feel of a rock, smell of a tree" [1]. So I started to think, hmmm, could we redraw them, shift them from circles to sponges, lines to waves, arrows to swells, triangles to rocks, and centres to tangles of stretching seaweed?

Sam: You asked me to find my favourites. That was quite a mission.

Ruth: Then I picked one to start with: D'Amato's figure 2 [3] (top right).

Sam: This is a great one to start with, it is about framing narratives, so in this first article we get a triple message: the insights from the diagram – both computing and creative perspectives – and the meta-observations from our journey.

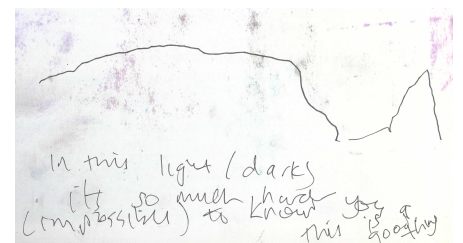
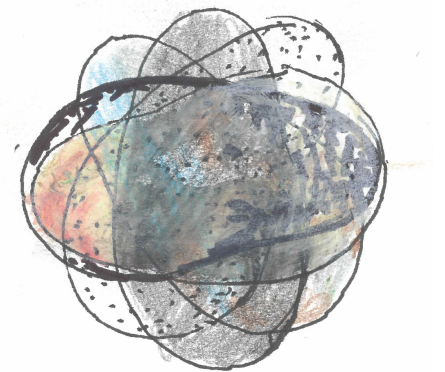
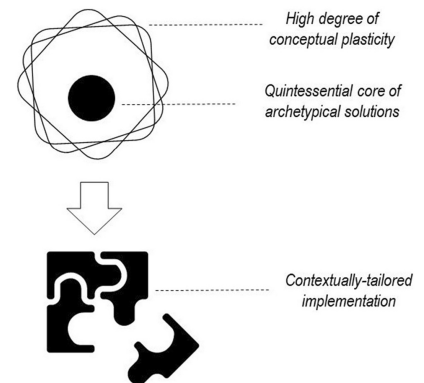
Ruth: I didn't read the paper until now, and we haven't discussed it, so your take is new to me.

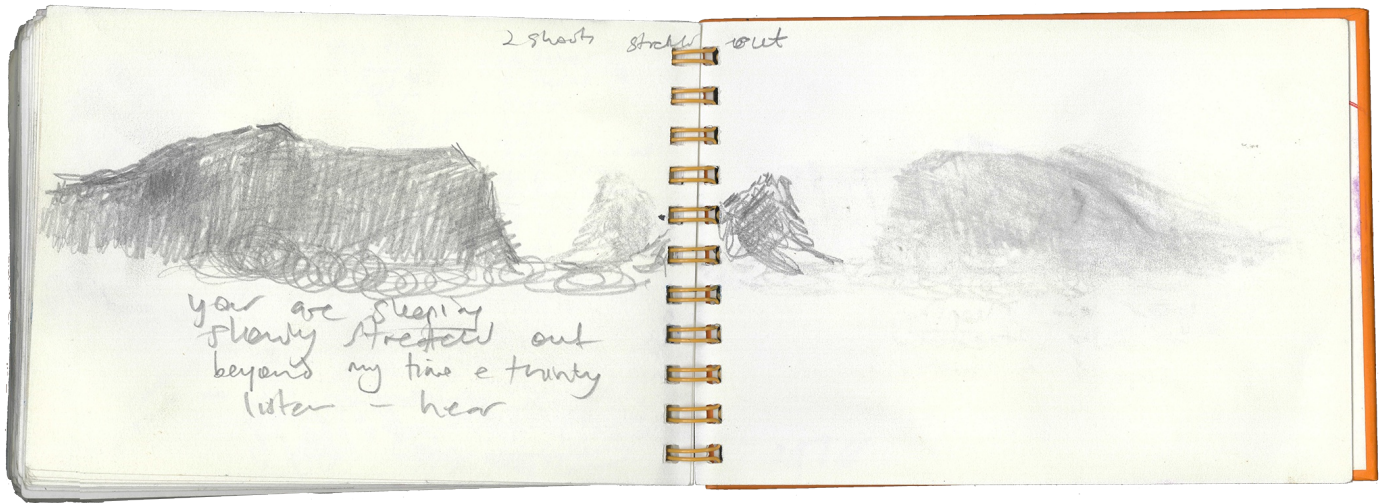
Sam: Dalia D'Amato [3] presents the circular economy as one of several sustainability narratives. She describes sustainability science as "normative, transdisciplinary, systemic and ontologically rich, encompassing all dimensions of human and (to a certain extent) nonhuman life and well-being". Changes in sociotechnical systems she describes as transformations. Trajectories in both emerge from political and discursive struggles that play out in complex, dynamic, and contested situations in ways that are highly contextual, often wicked and super-wicked problems. Consequently "there is no one-fits-all solution, but an open-ended and dynamic multiplicity of competing pathways".

Allowing for a diversity of potential pathways and solutions guarantees the inclusion of values and needs of multiple actors. Moreover, the possibility of choosing between multiple options improves the ability to respond to uncertainties and unexpected events, thus increasing adaptation and resilience capacities of environment-human systems. To be effective, interventions underpinning solutions should address multiple problems and their ultimate causes, be collective and collaborative, as well as engaging and intrinsically motivated, or in other words, intervention should be based on deep leverage points [3].

A narrative is presented as a generic way to describe a recurrent storyline, so a sustainability storyline is one used across sectors and in different systems.

She then introduces the notion of the sustainability narrative as pathways of change "used to identify macro-concepts that frame and address





one or multiple sustainability problems and offer a distinctive set of transferable and scalable solutions”. The circular economy is one such narrative, as is the sharing economy. Then comes

Ruth: Then comes the diagram...?

Sam: Yes, the diagram is then used to characterise sustainability narratives. The diagram has three primary elements - four if you count the arrow which isn't described but I think represents the process of getting from understanding to implementation. The top half is a set of overlapping rotationally offset rounded squares surrounding a small circle. These represent “a high degree of conceptual plasticity across stakeholders” (the squares), and a “quintessential core of archetypal solutions”. The lower part of the diagram is a black four-piece jigsaw puzzle, with one piece ajar, all representing a “tailored implementation leveraging and addressing contextual characteristics and needs”. These implementations combine with other narratives to “contribute to shaping pathways of change in socio-technical and social-ecological systems”.

Ruth: I really enjoy the diagram's fluidity.

Sam: The paper goes on to discuss the implications of framing the circular economy as a sustainability narrative. In short, the circular economy has conceptual plasticity (it is contested, diverse, multidimensional, wicked, overlapping complex systems); has a core of archetypal solutions (not singular solutions but embracing systems of change); and, needs tailored implementations (leveraging and addressing context across scales of territories, cultures, ecologies and so on). They conclude that socioecological and sociocultural change must be complementary to technology, that implementors of circular economies must pay attention to context, and solutions must be complementary and synergistic.

Ruth: Has anyone in computing noticed?

Sam: No, two of the 51 citations of D'Amato are from computing [9,12], but neither pick up on the diagram or the concept of sustainability narratives.

Ruth: So what do you make of it? Sam's Seven insights?

Sam: Seven? that's hopeful. OK then. You'd better keep count. I love this paper, not for the circular economy material but for the framing and reframing device represented by the diagram. So, the first observation is the notion of reframing- that we can stop and ask ourselves if what we are doing at the larger scale is the right thing to do. Second, is the prominence of narrative in that. The story we tell ourselves as computer scientists about what we are doing but also allowing for the possibilities and power of other stories [6, 15].

Third is the role of sociotechnical transitions. We

need to see that this is what we are doing in our computing, and see how that is providing leverage for socioecological change – all else is pointless.

Ruth: Bit strong?

Sam: Deliberately. 2023 will be the hottest year on record. “Our societies face clear and potent dangers that require urgent and transformative actions to protect present and future generations” [14]. It is urgent and will not work if not in tandem with socioecological change.

Fourth...computing can be seen as a grand socio-technical transition where computing represents modernism. While technological advances in the last centuries have seen improvements in standard of living, they have come with a huge cost. In describing Planetary Health, Whitmee et al. [14] emphasised the interdependent relationship between human health and the health of the natural world to the extent that improvements in human condition that occur at the expense of the environment are not considered to be progress. This, if nothing else, means that every computing project is a sustainability project.

Another point comes from sustainability narrative as a macro concept. We have adopted neoliberal efficiency as our narrative, without really stopping to see if that is only narrative. I've written before about the dangers of this [8].

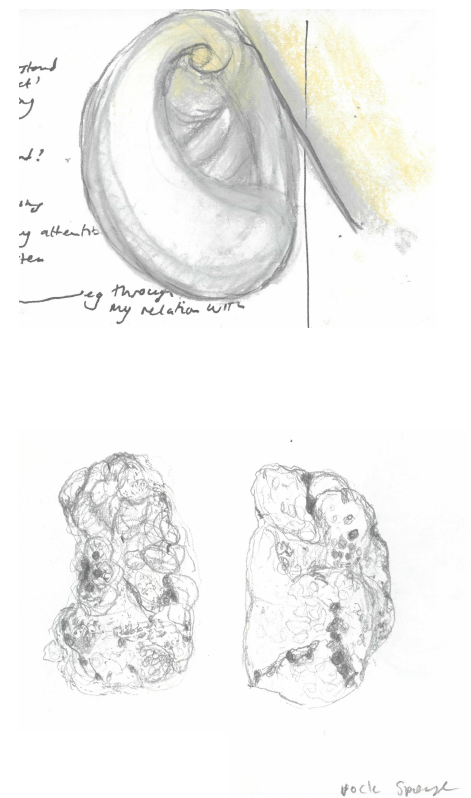
Ruth: Does the diagram offer any alternatives?

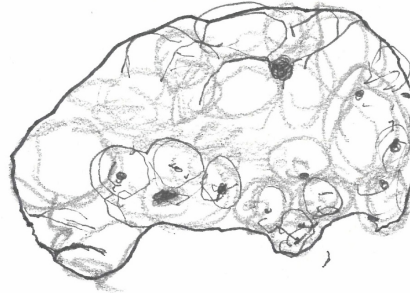
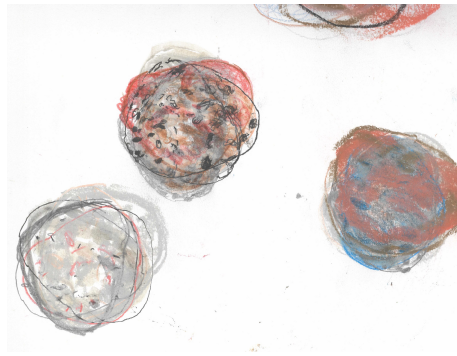
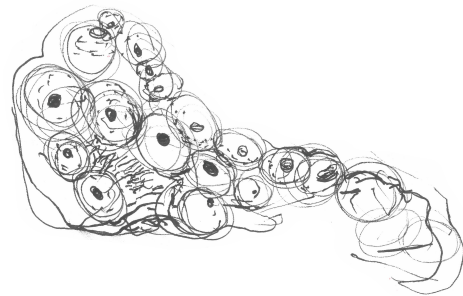
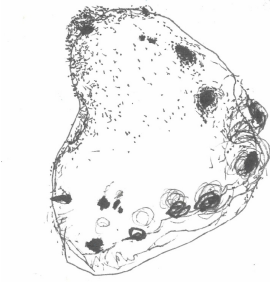
Sam: The idea of the quintessential core leading to contextual application aligns with Schuler's Liberating Voices patterning language [11].

Ruth: We've used those cards in teaching, you gave the cards to new post grads and had them pick out the cards that resonated, then helped them design their professional practice research around those. I think my favourites are Power of Story, Thinking Communities, Arts of Resistance and Appropriating Technology.

Sam: The plasticity of the problem and the multiplicity of core solutions mean sustainability is resistant to requirements engineering. The need for diverse voices, privileging the underrepresented, and the recognition of the wickedness of the challenge should mean a shift from imagining solutions to fix society and the environment as technical problems to nurturing moves towards ways of rethinking thriving futures. We still need user stories or requirements, but these need to focus on involving those people as part of the sociotechnical systems, not merely users. There are further opportunities if we think of intergenerational users or more-than-human stakeholders. I think that's six...

So, seven, and this follows the last one, we need to move to not imposing predetermined technical





I realise
I am
trying
to keep
staying
'not
understanding
You'

The sponge
Will help me
Soften the rock
The incredible complexities, holes, threads,
cones, rolls, bristling, folding and rolling into each
other,
clustering, clumping, opening
Listening to rock
Hold me, 'holding'

Rocks are
Porous like
Sponge
We are all
Porous
We are all
Holes, ears

Seeing, paying attention
Forming and reforming
Each of these bodies

Attach to things, other than me
She is calling us / out

solutions into contexts that require nuance. In the context of decolonising computing, Karetai et al. [6] characterise the unstated mindset of technological solutionism as “if technology doesn’t make things better for you, then there is something wrong with you, your education and your cultural understandings, not the technology”. We could extend that here to “if computing doesn’t fix your sustainability problem, then there is something wrong with your world?”.

What did you do?

Ruth: Well, I have been on a journey, mainly through connecting with place, which has been shifting and shaping my thinking in the world, especially around interconnectedness. So, I took the diagram to the beach with me. I continued drawing the islands, and I wondered. Then one day, diagram in hand and brushing past flaxes as I left the beach, I realised – within interrelatedness, there is something significant about the ‘I’ because it places us in it e.g., ethical choice, daily practice learning from mistakes, experiences. I’m sure D’Amato’s diagram helped me recognise this, make this connection.

My starting point then became somewhere around attachment, e.g., to feel, sense understand and potentially act, we might, I wonder first need a sense of attachment, connections, a knowing about, and develop understanding. This understanding is complex, looks like – fluid, porous, political, implicated, lived. I do this, with the diagram and place, through paying attention, listening, through my relations with rock, sponge and island.

Butler talking on Climate Sorrow [4,2], reminds us of our shared reality (through Winnicott’s self as relational, within which we exist and recognise ‘an other than me’), and extends this to encompass the air and soil. She explores her own ‘exchange relation’ with photographs of trees, recognising ‘without objects I can not be’. This attachment to thingness, my words as way of thinking through this, provokes a deeper understanding of a shared world.

Sam: Talking about thingness, that tiny shell I found

at Moeraki, the one I think looks like a baby shark helped you in this?

Ruth: Yes, what I saw in that tiny shell was straightaway an ear for me. This seemed to settle things, it was like listen, we can already hear the solutions, we can already find out a lot by paying attention to what is available. The shell ear pointed me back to Le Guin’s Carrier Bag Theory of Fiction [7], back to my thinking about the world as a sponge, which for me is about needing to soften to listen. Carrier Bags remind us to consider the stories we carry and how we carry them, this is important because this implicates where and what meaning can be found. As I drew, I listened to David Naïmon Crafting with Ursula [10], this polyvocality, became a gathering- with the sponge and stone, and the island, and the tiny shell too, all carrier bags. So, that tiny shell-ear helped me move - like the arrow in the diagram.

Sam: So then what?

Ruth: Well as I continued drawing sponge, rock, islands, holding the rocky-sponge pairings in my hand, sitting on the hill and the sand, leaning on the rocks, watching the islands, these activities spoke to each other, through each other. As I drew the sponge I became immersed in the holes, threads, bristles, cones, rolls, folding and rolling into each other, clustering, clumping around openings. And as I drew the stone, I saw similar openings, similar clumping and holding together, and then I would be back at the beach, drawing the islands, sea lions and birds calling, waves crashing, hill shuddering, colour complexities I couldn’t name. I was reminded again of the impossibility of these tasks - to know these things.

Sam: You made many drawings?

Ruth: Yes, the repeated attention is important – The other day my son was talking to me about how he wrote the word science over and over to cover the front of his science book, as he is writing, he starts to change the spelling, he misspells it. We were talking about what happens when you do that, how it makes you start to think more about the meaning of the

word. And this relates here, drawing the same thing over and over and with an aim to ‘not understand’ is a similar kind of process. It is to get in beneath the meanings that we may bring to it already, to develop new understandings and potentially think differently. This relates to Judith Butlers performativity [2] in that, if we think about performative utterance, as carrying meaning, we also recognise that such performance, may, significantly, be done differently.

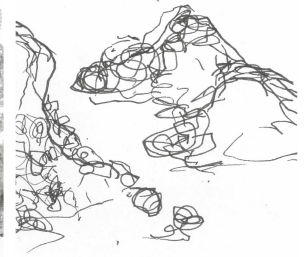
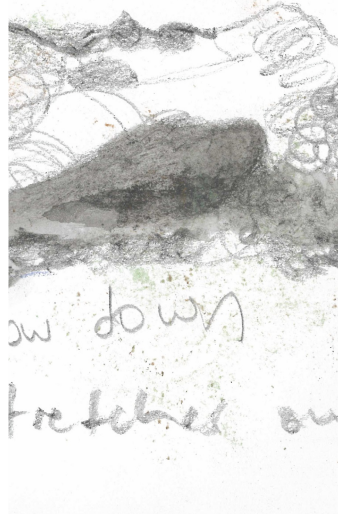
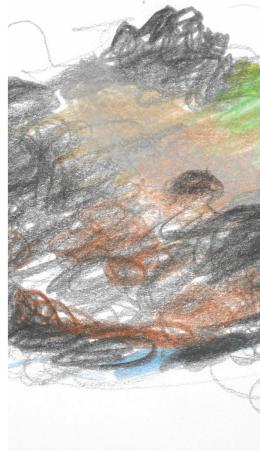
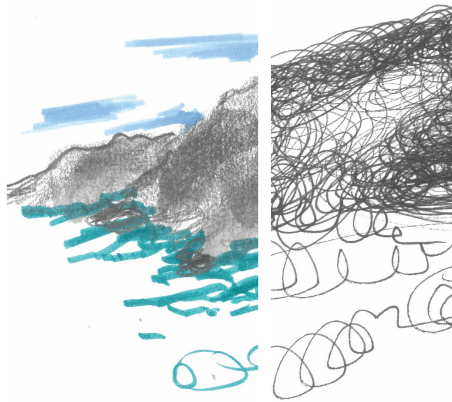
This returns us to Brady, these embodied, sensual practices are significant. They, like the repeated word of science, with hand, body, eyes, and physical accumulation of text, are the very world we live in and the problems we need to be considering, situated in. So if we can help ourselves to remember these locations, the felt, emotional, physical embodied significance of our thinking, and of these issues, perhaps we deepen or enrich our understandings?

Sam: What do you think we can take from your artistic explorations as insights for computing? Can we start with insights that stem from engagement with D’Amato’s diagram, and then some from the exercise as whole?

Ruth: Good plan. We need to be connected to matter. We need to suspend what we know in order to find ways to get at deeper learning – and this comes through allowing matter to help us connect on that level. This materiality might be a way to lift sustainability from being abstract and difficult to take notice of, it might help us increase the thingness of sustainability, and more fully understand our shared relationality? The materiality also means we engage with perspectives other than self, paying attention to Le Guin’s carrier bags. The rock and sponge are in between space, and to engage with them means a different form of collaboration, becoming available to other perspectives.

Sam: So, a deeper connectedness, more-than-human as process rather than an occasional stakeholder?

Ruth: Yes. We also need to be comfortable with ambiguity. Is it a rock, is it a sponge? Is it relevant? They merge and realising the holes in one makes me



look again at the other.

Sam: Are you saying that computing should be comfortable with not knowing?

Ruth: We need to be prepared to not look for an answer, a single answer, or even answers, maybe computing could be supporting us to look for questions.

Sam: And perhaps avoiding taxonomies or databases that force categorisation.

Ruth: Sometimes the detail is not where you started looking. The line drawing of the edge of the island – I wasn't looking for that, it was nearly dark, I could hardly see it.

Sam: The essence might not be in the obvious...

Ruth: ...nor in the measured or even the measurable. When I made the drawings of the sponge/rock that resemble the top of the diagram, I was repositioning the rock, turning about the middle – so multiple outlines. But that's not how you saw them.

Sam: I thought you had drawn the contours, horizontal slices.

Ruth: There are multiple ways of looking, representing and understanding.

Sam: All of which are right.

Ruth: All of which are not wrong. And can change over time. When I looked at the sponge, then the rocks I started to focus on the holes. We're holes as well. Then coming back to the diagram the black circle of the quintessential core became a hole but a hole of possibilities and kindness, and that placed me in the world. My practice, my ethics are all connected through the hole to those possibilities.

Sam: How do we as computer scientists connect to kind possibilities and to not close but open solutions?

Ruth: The arrow, it wasn't labelled on the diagram, but it drew me in, it connected the rock and the island. What is the relationship between the rock and the island? How is the whole beach connected in relationships, and how do they change?

Sam: How can we better support being relational and dynamic rather than ossified systems?

Ruth: Like my son's science book, I wrote "Stone observes slowly" over and over. I was transported to a different sense of time: Stone Time. In that time we need to look differently, be prepared to get lost and understand less. But it gives new insight. The islands soften over time – both my drawings and the islands themselves.

Sam: How can we make computing systems that default to geological time, or at least intergenerational processes?

Ruth: But the rock is also immediate. This relational thingness again. It is an instance of now. And my connection is bounded by the light and the waves and the wind – together as a temporal being.

Sam: Intergeneration and immediate.

Ruth: The rock time still exists in our time. We're still accountable. Maybe it's a humble ignorance [13].

Sam: It certainly tells computing that we need to find ways to deal with time differently. From a sustainability perspective we need to recognise that it isn't a series of short term, then a medium-term project then a long term goal. The long term has already started.

Ruth: We need to be comfortable with not understanding. We learn and progress from continuing disorienting and shifting.

Sam: That fits with D'Amato, I think the message for computing is need to stop focusing on the idea of there being a singular end-point. How can we make computer systems that enable this disorientation.

Ruth: There's something for computing too in biological metaphors. But not the ones that try to copy biology – not a biological look-alike.

Sam: Think like a rock...

Ruth: ...be a rock.

Sam: What did we get about the process of engaging in computing for sustainability through diagrams in a creative process? We've got a depth of insights here that we didn't get from just looking at the diagram.

Ruth: I'm reluctant to call it a process. It's a way of being. A way of approaching complexity. A way of engaging deeper. The diagram led me to looking at some stones. The shapes of the diagram had me look differently at the stones, helped me think relationally with them. Being outside my discipline – both sustainability and computing. I've worked with stone before as a sculptor, but here I found understandings being able to float through. This diagram, it's message about plasticity, multiple cores, and implications in different contexts. I think there's two messages here – be comfortable being uncomfortable outside your discipline. Go and talk with some stones, or at least some people in fields you have no idea about. And don't afraid to bring the deeper emotional engagement to computing.

Sam: Do you think computing has protected itself by its abstraction?

Ruth: Of separating itself from the materiality of the world? Yes.

Understanding takes dwelling. Allow for emergence. Perhaps also we may be missing out on embodied understandings, at the level of meaning, for instance performative missteps.

Sam: How might we allow for that in computing? Perhaps the contingency at the end of the project could be supplemented with time at the start to explore.

Ruth: Yes, allow for ruptures. Want the process to fundamentally shake up you way of doing things. One such rupture would be a focus of how can computing adopt a kind relational humbleness rather than a mindset of technical solutionism? Another is getting beyond the drive to code it, simplify and make efficient, which seems to me to reinforce problems.

Sam: It is interesting to me that the diagram we started with yielded interesting insights for computing, probably beyond what the author intended, but then this process has pushed beyond the limits of what computing can do, and hopefully given people something to think about. Let's do it again next issue. Jochem's diagram of planetary health literacy [5].

Ruth: It's a globe, a funnel, cyclic processes, people, scales...back to the beach I think...

References

- [1] Ivan Brady. 2004. In Defense of the Sensual: Meaning Construction in Ethnography and Poetics. *Qualitative Inquiry* 10, 4 (2004), 622–644. <https://doi.org/10.1177/107780040426>
- [2] Judith Butler. 1988. Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal* 40, 4 (1988), 519–531. <https://doi.org/10.2307/3207893>
- [3] D. D'Amato. 2021. Sustainability Narratives as Transformative Solution Pathways: Zooming in on the Circular Economy. *Circular Economy and Sustainability* 1, 1 (2021), 231–242. <https://doi.org/10.1007/s43615-021-00008-1>
- [4] Undisciplined Environments. 2023. Judith Butler on climate sorrow. <https://undisciplinedenvironments.org/2023/06/26/>

judith-butler-on-climate-sorrow/

- [5] Carmen Jochem, Julia von Sommoggy, Anna-Katharina Hornidge, Eva-Maria Schwienhorst-Stich, and Christian Apfelbacher. 2023. Planetary health literacy: A conceptual model. *Frontiers in Public Health* 10 (2023), 980779. <https://doi.org/10.3389/fpubh.2022.980779>
- [6] Mawera Karetai, Samuel Mann, Dhammika Dave Guruge, Sherlock Licorish, and Alison Clear. [n. d.]. Decolonising Computer Science Education-A Global Perspective. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 1097–1102. <https://doi.org/10.1145/3545945.3569870>
- [7] Ursula K. Le Guin and Bul Lee. 2019. *The carrier bag theory of fiction*. Ignota Books, London, England. Includes bibliographical references.
- [8] Samuel Mann and Oliver Bates. 2017. A conversation between two sustainable HCI researchers: The role of HCI in a positive socio-ecological transformation. Routledge, Book section 3, 45–60. https://doi.org/10.9774/leaf.9781315465975_7
- [9] Florin Mihai, Ofelia Ema Aleca, Andrei Stanciu, Mirela Gheorghe, and Mirela Stan. 2022. Digitalization - The Engine of Sustainability in the Energy Industry. *Energies* 15, 6 (2022), 2164. <https://doi.org/10.3390/en15062164>
- [10] David Naimon. [n. d.]. *Crafting with Ursula Archives*. https://tinhouse.com/th_podcast_cat/crafting-with-ursula/
- [11] D. Schuler. 2008. *Liberating Voices: A Pattern Language for Communication Revolution*. MIT Press. <https://doi.org/10.7551/mitpress/7627.001.0001>
- [12] Eirini Triantafyllidou and Anastasia Zabaniotou. 2021. Digital technology and social innovation promoting a green citizenship: development of the “go sustain-able living” digital application. *Circular Economy and Sustainability* (2021), 1–24. <https://doi.org/10.1007/s43615-021-00111-3>
- [13] B. Vitek and W. Jackson. 2008. *The virtues of ignorance: complexity, sustainability, and the limits of knowledge*. University Press of Kentucky, Lexington. 354 pages. <https://doi.org/10.4198.118>
- [14] Sarah Whitmee, Andy Haines, Chris Beyrer, Frederick Boltz, Anthony G Capon, Bráulio Ferreira de Souza Dias, Alex Ezech, Howard Frumkin, Peng Gong, and Peter Head. 2015. Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health. *The Lancet* 386, 10007 (2015), 1973–2028. [https://doi.org/10.1016/S0140-6736\(15\)60901-1](https://doi.org/10.1016/S0140-6736(15)60901-1)
- [15] Marisol Wong-Villacres, Cat Kutay, Shaimaa Lazem, Nova Ahmed, Cristina Abad, Cesar Collazos, Shady Elbassuoni, Farzana Islam, Deepa Singh, Tasmiah Tahsin Mayeesha, Martin Mabeifam Ujakpa, Tariq Zaman, and Nicola J. Bidwell. 2023. Making Ethics at Home in Global CS Education: Provoking Stories from the Souths. *ACM J. Comput. Sustain. Soc.* (aug 2023). <https://doi.org/10.1145/3608113>

Samuel Mann

Ruth Myers

Capable NZ, Te Pūkenga,
Dunedin, New Zealand

samuel.mann,ruth.myers@op.ac.nz

NEWS



Photo: © Johanna Blumenthal



Third Annual SIGCAS Showcase



After two successful Showcases in 2021 and 2022, the current Executive Committee is currently discussing approaches to ensure the third SIGCAS Showcase best serves the needs of our SIG. Previously, the Showcase was conducted as two-half day sessions on two consecutive days. Potential new approaches include a one half-day event and one-hour recorded sessions spread throughout the year.

In case you haven't been able to attend the early events, the virtual Showcase event included:

- Presentation of SIGCAS awards
- Keynote Address

Multiple Panels with renowned panelists:
SIGCAS Town Hall

We encourage the entire SIGCAS community to participate in this event and are soliciting suggestions for how this Showcase event can best serve the needs of our members. You can submit ideas by speaking with one of the Executive Committee members or sending the suggestion to editors sigcas@acm.org.

— continued from page 8 —

I feel that the take-away here is to encourage ethical consideration at an earlier stage of computer learning; to ensure students are internalizing the idea of the wide range of stakeholders that count on their willingness to ensure that “the way it is” can always get better.

- [1] Thomas Davenport, Ravi Kalakota. 2019. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*. (June 6, 2019). DOI: 10.7861/futurehosp.6-2-94
- [2] Discover Ways to Collaborate. IBM Developer Site. Retrieved Oct. 26, 2023 from https://developer.ibm.com/?utm_content=SRCWW&p1=Search&p4=43700074472741094&p5=e&gclid=CjwKCAjwnOipBhBQEiwACyGLukuNPjs9sX4lv-bPbLS-U9kk1IEi5qC6WqH45qOuE3hN_Dh66Uh-BoCoKcQAvD_BwE&gclsrc=aw.ds
- [3] Junior Year Writing Course Criteria. UMass Amherst. Retrieved Oct. 26, 2023 from <https://www.umass.edu/writingprogram/junior-year-writing-requirements>
- [4] Sara Reardon. 2023. AI Chatbots Can Diagnose Medical Conditions at Home. How Good Are

They? *Scientific American*. (March 31, 2023) Retrieved Oct. 26, 2023 from <https://www.scientificamerican.com/article/ai-chatbots-can-diagnose-medical-conditions-at-home-how-good-are-they/>

- [5] Michael Rigby. 2019. Ethical Dimensions of Using Artificial Intelligence in Health Care. *AMA Journal of Ethics*. (Feb 2019) DOI: 10.1001/amajethics.2019.121.
- [6] David Robinson. 2022. The Kidney Transplant Algorithm’s Surprising Lessons for Ethical A.I. *Slate.com* (Aug. 31, 2022). Retrieved Oct. 26, 2023 from <https://slate.com/technology/2022/08/kidney-allocation-algorithm-ai-ethics.html>
- [7] Matthew Sweeney. 2022. Ethical dilemmas in computer science. *ZDNET*. (Feb 11, 2022) Retrieved Oct. 26, 2023 from <https://www.zdnet.com/education/computers-tech/ethical-dilemmas-computer-science/>

Christina Sutcliffe
UMass

Comparison of Rationales and Approaches of Artificial Intelligence Regulatory Policies in the U.S, EU and China Draws Recommendations on a ‘Fit for Purpose’ Global AI Governance Mechanism

By AO KONG, WEIYUE WU and SHAOSHAN LIU

Keywords: Artificial intelligence, regulatory policy
 Categories: Social and professional topics → Computing / Technology policy

Introduction

With the rapid development of AI technologies, our society is facing a global AI regulatory dilemma [1], with a wide spectrum of AI regulatory rules not only confusing but also incurring very high costs for the participants of the AI industry [2].

This paper concentrates on analyzing and comparing representative regulatory rationale and approaches of Artificial Intelligence (AI) from the United States (U.S.), the European Union (EU), and China, shedding light on their respective advantages and limitations and exploring their implications for the formation of a global AI regulatory mechanism.

Moreover, this paper outlines “fit for purpose” considerations for global AI regulation, ranging from variances in identifying priorities and accepting tradeoffs when deploying AI in different international development contexts, to the composition of the governing board. In this context, the International Atomic Energy Agency (IAEA) is juxtaposed as an example of existing international regulatory bodies governing high-impact and high-public-risk technology, to explore how AI’s highly self-generative and open-source nature necessitates a unique governance model.

Building on this, we propose a global open-source public goods governance mechanism for AI technology that upholds safety, dignity, and equity standards; ensures broad representation across geopolitical, technical, and socioeconomic spectra; aligns with national development priorities; and accommodates AI technology’s unique features. In this endeavor, the UN’s High-Level Advisory Board on AI could assume a pivotal convening role to shape such a mechanism and foster its contribution to the realization of Sustainable Development Goals.

2. Comparison of Rationales and Approaches of AI Regulatory Policies in the U.S., EU and China

2.1 Key Concepts

In the examination of various AI regulatory models, we introduce two key terms: **Rationale** refers to the guiding principle that dictates the focus of regulation. It sets the boundaries of regulatory intervention by identifying which aspects of AI technology are critical for scrutiny and control. **Approach** addresses the execution methods of these regulations. It looks at the structural design of the regulatory system, the assignment of roles among regulatory bodies, and the mechanisms through which rules are enforced. These two terms serve as a structured lens through which different AI regulatory

frameworks can be assessed and compared.

2.2 The U.S.: Industry-Specific Strategy

While the U.S. has yet to finalize a specific regulatory strategy for AI [3], analyzing legal frameworks pertinent to AI’s essential components and considering the current White House’s call for action provide insightful reflections. For relevant legal comparison, we dissect AI into its core components as follows:

a) **Data Protection:** We investigated the implementation of U.S. privacy laws related to data protection and confidentiality to understand the landscape of data safeguards.

b) **Algorithm Regulation:** Principles and methods leveraged in cybersecurity offer valuable insights for regulating algorithms, as they are essentially software requiring secure measures.

c) **User Protection:** In discerning protections for AI users, we explored consumer protection principles and practices, acknowledging that users of AI—being consumers of this emerging technology—have legal rights that necessitate protection.

Our research indicates that within each of the three legal domains—privacy, cybersecurity, and consumer protection—the U.S. consistently adopts an **industry-specific rationale**. Regulations in the U.S. are developed in a bottom-up manner, allowing various industry lobby groups to craft initial versions of regulatory laws, which are then iteratively refined through the legislative branches. Essentially, the legislative system in the U.S. grants each industry the autonomy to propose regulatory laws pertinent to its sector [4].

In terms of regulation enforcement, the U.S. approach is also industry-specific. For instance, cybersecurity is a vital concern across all industries, especially as businesses increasingly operate in digital spaces. Cybersecurity is overseen primarily by the Cybersecurity and Infrastructure Security Agency within the Department of Homeland Security. However, other agencies like the Federal Trade Commission (FTC) and the Securities and Exchange Commission (SEC) hold industry-specific responsibilities, such as penalizing companies that fail to protect consumer data in the case of the FTC, or regulating cybersecurity disclosures in the financial industry in the case of the SEC. Similarly, consumer protection is handled across multiple agencies and industries. While the FTC is the main consumer protection agency, other agencies, such as the Consumer Financial Protection Bureau and the Food and Drug Administration, play important roles within

specific industries.

A notable indication of the U.S.’s industry-specific approach is reflected in the White House’s request for voluntary commitments from leading AI companies to manage AI risks [5]. This request shows the government’s trust and reliance on these companies’ ability to effectively govern their own AI applications. A prime example is Meta, previously known as Facebook, which has established a Responsible AI team and launched the “Community Forum on Generative AI” to gather public feedback on AI products in a transparent manner [6].

Advocates of the U.S.’s industry-specific approach contend that expert panels, primarily composed of experienced industry practitioners, possess the most in-depth and thorough understanding of a particular sector [7]. These expert panels, already well-versed in industry nuances, can be further enhanced with the inclusion of AI specialists. By doing so, a well-informed and intricate regulatory framework can be developed for AI applications across a range of industries.

Significant concerns about this industry-driven approach include the arbitrary nature of self-regulation and the influence from monopolies or oligopolies when a handful of top companies wield dominant influence [8]. AI, with its transformative impact and fast adoption unlike any technology before, needs to be cognizant of the downside of mainly relying on “good-faith” and the potential dominance over norms set by a few sectors and industries.

2.3 The EU’s GDPR-aligned Strategy

The EU’s AI Act follows the framework established by its predecessor, the General Data Protection Regulation (GDPR) [9]. The Act proposes a comprehensive structure for AI regulation, spanning from defining requirements for high-risk AI systems to the establishment of a European Artificial Intelligence Board. The Act emphasizes user safety and fundamental rights, mandates transparency of AI systems, and enforces strict post-market monitoring rules for AI providers. This legislative initiative reflects the EU’s dedication to cultivating a human-centric and ethical AI ecosystem and to protecting the broader public in the realm of AI technology.

The rationale behind the EU’s AI Act is risk-based. It classifies AI products into distinct categories, with each being subject to different levels of regulatory requirements, assessing the potential harm an AI product could inflict and prescribing the necessary safeguards. Low-risk AI systems, like spam

	Rationale	Approach	Benefits	Limitations
U.S.	Industry-specific	Industry-specific	Preserves existing regulatory frameworks Leverages industry-specific expertise	Arbitrary nature of industry Risk of monopoly and oligopoly
EU	Risk-Based	Centralized	Offers a holistic oversight of AI regulation Sets clear and unified standards	Lack of adaptability and scalability
China	State-Controlled	State-Controlled	In line with National Development Plan and pace Prevent industry's overinfluence	Vulnerable to significant risks Not enable innovative breakthroughs

Table 1: Summary of AI Regulation Proposals

filters or video game algorithms, may face minimal regulatory constraints to maintain innovation and usability. In contrast, high-risk AI applications, such as those in biometric identification and critical infrastructure, are bound by extensive obligations, including stringent risk management and user transparency requirements [10].

To implement the Act, the EU's approach is to establish a centralized regulatory body, namely the Committee on Artificial Intelligence [11], responsible for detailing the legal framework for AI, interpreting, and enforcing the AI Act's regulations and supervising high-risk AI systems to ensure uniform application across the union.

While aiming to provide a cohesive regulatory environment, the implementation of the EU's AI Act might encounter obstacles similar to those of the GDPR, which faced criticism for its unintended consequences and complex rules that burdened companies and did not significantly enhance user trust or experience [12]. The risk-based rationale could prove overly simplistic for the complex realities of AI products, overlooking the inherent ambiguities and diverse risk scenarios associated with AI systems. A recent study [13] revealing that a significant number of AI systems could be categorized as high-risk highlights this potential flaw, suggesting that this approach might impose excessive regulatory burdens, hindering the development of beneficial technologies [14].

Given the fast-paced evolution of AI and its global deployment, a single, centralized regulatory entity, despite its comprehensive approach, may find it challenging to address the diverse and rapidly changing landscape of AI-related issues effectively. Potential decision-making bottlenecks and bureaucratic delays could hinder timely responses essential in the dynamic AI landscape, impacting **regulatory efficacy and adaptability**. While the intent behind establishing a centralized Committee on Artificial Intelligence is commendable, its practical effectiveness in navigating real-world complexities is yet to be determined.

2.4 China: State-Control in AI Regulation

The rationale and approach of China's AI regulations are mostly state controlled. China views AI as not just a technological advancement, but as an integral part of its economic and social infrastructure, much like the traditional public goods it has regulated, such as energy and electricity [17]. The primary objective behind is to foster a safe and measured pace of rolling out AI and its applications through state control [18] and to prevent undue dominance or monopolistic tendencies from the private sector.

The recent generative AI regulations demonstrate the authority's commitment to this approach. The regulatory framework resonates with the principles found in the Cybersecurity Law [15], extending similar

oversight responsibilities, previously earmarked for Internet Service Providers (ISPs) and social media platforms, to generative AI service providers. Just as ISPs are tasked with content oversight, AI service providers must operate under the supervision of regulators and report their maintained records to regulatory bodies [18]. The swift rollout of these regulations within just five months of the debut of ChatGPT illustrates China's regulatory authority's resolution to maintain its pace with AI's rapid evolution [19].

When executed in accordance with the National Development Plan, this state-controlled AI regulatory model could help ensure that the development and deployment of AI align with the country's developmental vision, strategic priorities, and pace. This is particularly compelling for developing countries, where there exists a need to guard against the rapid dissemination of AI technology and its impacts.

Meanwhile, unlike traditionally regulated public goods in China, such as land, mines, and electricity, the dynamic nature of AI necessitates a constantly updated knowledge base, agile regulatory frameworks, and significant investments in computing power. The challenge resides in striking the right balance—implementing a regulatory mechanism robust enough to safeguard public interest while still being flexible enough to encourage innovation and allow for industry experimentation.

The rationale, approach, benefits and limitations of each above-mentioned regulatory frameworks are summarized in the Table 1 below.

3. "Fit for Purpose" for Global AI Regulation: Essential Features of a Global Governance Mechanism

Since AI technology and its impacts are not confined by any borders, it is imperative for the UN Advisory Board on AI to contemplate a unified global mechanism for AI regulation that can bridge cultural and policy divides.

Constructing a global AI regulatory mechanism that is genuinely "fit for purpose" poses a formidable challenge. As depicted by the distinct regulatory approaches from the US, EU, and China, one of the paramount considerations is navigating the intricate web of socioeconomic and political differences and the deeply rooted regulatory traditions inherent in each nation's legal and administrative systems.

Countries also weigh the regulatory priorities and trade-offs of AI technology in their respective contexts. Advanced economies may prioritize risk mitigation and privacy protection, while developing nations might seek to leverage AI for economic growth and to address urgent societal challenges. Achieving equilibrium between these varied objectives mandates the UN to utilize its unique position to facilitate cross-cultural dialogue and diplomacy,

harmonizing conflicting perspectives.

AI's open-source and self-generative nature necessitate an agile, responsive governance mechanism, surpassing those in existence for other high-impact, high-risk technologies, like nuclear power. Some proponents suggest establishing an International AI Agency [20-22], conceptualized as a centralized repository for AI regulation akin to the role of the International Atomic Energy Agency (IAEA) in nuclear governance, to guide national AI strategies and bridge policy gaps as AI technology progresses.

However, we emphasize that the efficiency of the IAEA is attributable to the manageable number of nuclear entities it oversees, with approximately 440 nuclear power reactors worldwide and nuclear armament restricted to nine nations [23, 24]. Nuclear risks, given their catastrophic potential, have prompted the adoption of globally accepted stringent safety protocols. Conversely, AI, with its open-source nature and significant influence from non-state actors, may necessitate an open-source regulatory monitoring and agenda-setting platform, resembling "GitHub," a cloud-based service platform commonly used to host open source software development projects, rather than a traditional, centralized governance model with periodically convened consultative sessions.

AI's ubiquity across domains and its array of applications, from industrial to military use, coupled with its diverse and stark risks including massive job displacement, deep fakes, and weapons automatization necessitate a broad representation from diverse socioeconomic sectors, geographic contexts, and ethnic groups to ensure inclusive decision-making.

In summary, the evolution of AI is in its nascent stages, but its exponential and unmanaged growth is already forecasting a pandemic-like scenario without urgent and proper intervention now. Based on the aforementioned research and analyses, we suggest a global open-sourced public goods governance mechanism [25] that upholds safety, human dignity, and equity standards; ensures diverse representation across geopolitical, technical, and socioeconomic profiles; respects national priorities and cultural contexts; and responds to AI's self-generative and open source nature could lay a solid foundation for global AI regulation.

References

- [1] Wu, W. and Liu, S. (2023). Dilemma of the artificial intelligence regulatory landscape. *Communications of the ACM*, 66(9), pp.28-31.
- [2] Wu, W. and Liu, S. Why Compliance Costs of AI Commercialization Maybe Holding Start-Ups Back. *Kennedy School Review*
- [3] O-Shaughnessy, M. and Sheehan, M. (2023). "Lessons from world's two experiments in ai governance." *Carnegie Endowment International*. <https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai>

- governance-pub-89035
- [4] McGrath, C. & Harris, P. (2009). The Creation of the US Lobbying Industry. In D. Johnson (Ed.), *Handbook of Political Management* (pp. 407-419). Routledge.
- [5] The White House, “Fact sheet: Biden-Harris administration secures voluntary commitments from leading artificial intelligence companies to manage the risks posed by ai.” <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>. Accessed: 2023-07-20.
- [6] Clegg, N. Bringing People Together to Inform Decision-Making on Generative AI. *Meta*. <https://about.fb.com/news/2023/06/generative-ai-community-forum/>. Accessed: 2023-09-18
- [7] Montgomery, C. (2023). Why a single ai regulator just won't work. *The Hill*. <https://thehill.com/opinion/congress-blog/4020159-why-a-single-ai-regulator-just-wont-work/>. Accessed: 2023-07-05.
- [8] Foge, Z. “American oligarchy: How the enfeebling of antitrust law corrodes the republic,” *J. Bus. Entrepreneurship & L.*, 12, p. 119, 2019.
- [9] European Parliament. (2023). AI Act: a step closer to the first rules on artificial intelligence. https://www.europarl.europa.eu/pdfs/news/expert/2023/5/press_release/20230505IPR84904/20230505IPR84904_en.pdf. Accessed: 2023-07-05.
- [10] EU commission (2023). Regulatory framework proposal on artificial intelligence. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>. Accessed: 2023-07-05.
- [11] Council of Europe. (2023). The council of Europe & Artificial Intelligence.” <https://rm.coe.int/brochure-artificial-intelligence-en-march-2023-print/1680aab8e6>. Accessed: 2023-07-05.
- [12] Chivot, E., Castro, D. (2019). “What the evidence shows about the impact of the GDPR after one year,” *Center for Data Innovation*.
- [13] AAI (2023). AI Act: Risk classification of ai systems from a practical perspective.” <https://aai.frb.io/assets/files/AI-Act-Risk-Classification-Study-appliedAI-March-2023.pdf>. Accessed: 2023-07-05.
- [14] European Commission (2023). Impact assessment of the regulation on artificial intelligence.” <https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence>. Accessed: 2023-07-05.
- [15] Liu, S. (2023). “Artificial intelligence will bring social changes in china,” *the Diplomat*.
- [16] Liu, S. (2023). “Will china create a new state-owned enterprise to monopolize artificial intelligence?”, *the Diplomat*, 2023.
- [17] Carnegie Endowment. (2023). China's AI regulations and how they get made.” <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>
- [18] Measures for the management of generative artificial intelligence services.” http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.
- [19] Digichina. (2017). Translation: Cybersecurity law of the people's republic of china (effective June 1, 2017). Stanford University. <https://digichina.stanford.edu/work/translation-cybersecurity-law-of-the-peoples-republic-of-china-effective-june-1-2017/>. Accessed: 2023-07-05.
- [20] Heath, R. and Gold, A. (2023). AI leaders: Sam Altman and others call for regulation in senate testimony. Axios <https://www.axios.com/2023/05/17/ai-leaders-sam-altman-regulate-senate>. Accessed: 2023-07-05.
- [21] United Nations. (2023). Secretary-general urges broad engagement from all stakeholders towards united nations code of conduct for information integrity on digital platforms. <https://press.un.org/en/2023/sgsm21832.doc.htm>. Accessed: 2023-07-05.
- [22] The Economist. (2023). The world needs an international agency for artificial intelligence, say two ai experts. <https://www.economist.com/by-invitation/2023/04/18/the-world-needs-an-international-agency-for-artificial-intelligence-say-two-ai-experts>. Accessed: 2023-07-05.
- [23] WNA (2023). Plans for new reactors worldwide. World Nuclear Association. <https://world-nuclear.org/information-library/current-and-future-generation/plans-for-new-reactors-worldwide.aspx>. Accessed: 2023-07-05.
- [24] Kristensen, H, Korda, M., Johns, E. and Kohn, K. (2023). Status of world nuclear forces. *Federation of American Scientists*. <https://fas.org/initiative/status-world-nuclear-forces/>. Accessed: 2023-07-05.
- [25] Wu, W. and Liu, S. (2023). The Need for a Global Standard of Artificial Intelligence Regulation. *ACM SIGCAS Computers and Society*, 52(1), pp.16-16.

Ao Kongyue Wu
UN Resident Coordinator System
kongyu@un.org

Weiyue Wu
Multi-National Corporation Compliance Auditor,
KPMG
Senior Automobile Consultant,
Deloitte
IEEE Special Technical Community on
Autonomous Driving Technologies
weiyue.wu@perceptim.io

Shaoshan Liu
Asia Chair, IEEE Entrepreneurship,
Member, ACM U.S. Technology Policy Committee
shaoshan.liu@perceptim.io

SHORT PIECE

Creative Coding as Method and Goal in a Gender Diverse First-Year Seminar

BY PAIGE GULLEY

FIERCE (Fostering Inclusion by Engaging in Real-world Computing Education) is a first-year academic program for gender diverse students entering UMass's computer science or informatics degree program. In response to feedback that students wanted to be exposed to more actual programming prompts in addition to the more theoretical content which was the focus of the previous year's seminar, this year's seminar is focused on Creative Coding.

The pedagogical goal of this class is to develop student's relationship to the craft of programming. Especially given the current heat around generative AI, it feels important to give students the opportunity to learn that writing code can be rewarding in and of itself. By the same token, as we are in the midst of adapting to this new paradigm, it feels especially important to teach students how to analyze the impact of technology on their own lives and on the world. Because creative coding is necessarily open-ended and driven by curiosity, it's an ideal medium for exploring this space.

Using Processing.js, a web-native implementation of the popular Processing framework, students are tasked with both ideation and execution in response

to readings, lectures, and creative prompts. The output of each assignment is a javascript application (or "sketch") which draws a dynamic image to a web canvas. p5js.com provides a simple cross-platform way to author and share these sketches. Because these are first year students with little if any prior programming experience it is necessary to provide boilerplate code and to encourage them to work in groups. The hope is that all students' will have the opportunity to engage in the creative process of conceptualizing their response to prompts, even if they don't have their hand on the wheel of executing it. Critical feminist content is still present, serving as tone setting side-dishes around these prompts.

A successful recent assignment asked students to consider beauty and epistemic biases. Prior to receiving this assignment they read "Asters and Goldenrod," a chapter from Braiding Sweetgrass by Robin Wall Kimmerer in which she relays her experience as an indigenous woman entering a college botany program, being derided when she explicated interest in botany as "wanting to know why asters and goldenrod look so beautiful together." Students were asked to take a picture of something on campus they

found beautiful.

The creative coding assignment itself asked students to write code which would "annotate" their images in a way which drew attention to what they found beautiful in them. The results were diverse and encouraging- one group of students who took a picture of a koi pond on campus implemented a ripple effect, which radiates out from the mouse when the image is clicked. Another pair of students who took a picture of an early autumn sunset produced a sketch which frames the image in a swirling gradient of colors selected from the image.

Positive responses to this structure and approach to teaching coding concepts has been incredibly encouraging, and I am eager to observe future student output. Although development of this course is still in progress, early indications are that it is successfully engaging students' creativity and critical thought, and that students seem to genuinely appreciate having these experiences.

Paige Gulley
University of Massachusetts
pgulley@umass.edu

Project Unplug – A Personal Reflection on Rediscovering the Joy in Being Online

BY MICHAEL JAMES HERON

In 2021 I realised that large swathes of the Internet were making me sad. I resolved to do something about it.

The result was a focused, intense process of finding the mechanisms by which I could optimise my relationship with the online world. I resolved to focus on those things that brought me sustained happiness, as opposed to those things that merely offered transient pleasure. I ripped toxic websites out of my life, and limited my exposure to those I regarded as merely hazardous. The result was a battle to regain joy I once experienced in my online leisure activities – to regain my focus on electronic common ground, as opposed to digital battle grounds.

Few technologies have reached so deeply into our lives – and into our sense of self – as the modern Internet. Social and psychological defence mechanisms that evolved for the context of a tribe simply do not survive at Internet scale. For all that social media is designed to link us together, in many ways it only serves to push us apart.

This paper is a personal, autoethnographic reflection of how ‘Project Unplug’ was conceptualised, prepared, and executed over 2022. The specific combination of online vectors that were the cause of my unhappiness are unique to me.

This paper though outlines a hopefully valuable approach to creating our own digital rituals and rites so as to create strong, inviolable limits on the extent to which the Internet might erode our personal happiness

Keywords: Autoethnography, internet usage, surveillance capitalism, social media
Categories: Applied computing → Sociology, Ethnograph

Introduction

In 2020 I made a monumental personal discovery. It was this - a lot of the time I spent on the Internet was making me unhappy. I'd sit down in a perfectly agreeable mood, and within a few minutes of opening up Reddit I was listlessly scrolling as a growing sense of misery gnawed away at me. I'd stand up in a foul mood which would slowly fade away in proportion to the time I was absent from a screen. As someone who, then and now, spends a ridiculous percentage of his life in front of a screen, it was a problematic epiphany. And I know I am not alone. The mental health and life satisfaction declines associated with Internet usage are well discussed in the literature [18, 26, 27] and while conclusions along vary with survey style and ideological slant, the connection is troubling. Social media platforms have been linked to everything from self-image problems in teenagers [6, 31] to a disengagement with critical thinking [5, 10]. It has been argued by some, such as Haidt and Allen [11] and Twenge et al. [28] that societal maladies, particularly amongst the young, can be linked directly to the always-available nature of smartphones [28]. Our relationship to the Internet is what enables much of the infrastructure of modern society. However, it increasingly feels to me like it may be the modern incarnation of Kurt Vonnegut's quote - "Future generations will look back on TV as the lead in the water pipes that slowly drove the Romans mad". Here I do not argue for the validity of any of these perspectives, but rather they reflect how I feel about my own relationship to always-on, omnipresent connectivity. In that respect, they have an emotional authenticity that reflects my own biases.

I mulled on all of this for some time over 2020. In 2021, I first expressed where my thoughts were going, ironically, on Facebook. On the 6th of March, 2021 I posted a status update, which included the following:

Realistically, is it possible to unplug from the internet these days?

I'm pondering this as a challenge to myself in 2022 and I'm just not sure it can be done while also holding down a job like mine. I really want to do it though. I think my happiness is inversely proportional to my Internet usage.

Anyone had any thoughts about doing this before? Anyone made any progress on a plan? Anyone else

tempted?

For a long time now, at the dawn of a new year, I set myself a personal growth project. The most significant of these was in 2019 where I undertook a 'depth year' - a year in which I bought no new things and started no new hobbies. Other years have involved taming my backlog of unplayed games (a project which, as of the time of writing, has taken over 3200 focused hours) and hitting an array of personal reading goals. This Facebook post marked the start of the preparation for what I termed Project Unplug 2022, in which I would attempt to craft the optimal online experience. I don't know if it's possible to have spoilers in an academic paper, but if so - spoiler alert.

I began by asking if it was possible to unplug myself from the Internet. The answer there is a resounding no. Given my professional responsibilities, it's just not feasible. On top of this I live in Sweden, and given the extent to which much of the country's infrastructure is digitised, it's also not feasible. I believed that it was certainly possible to have a much more pleasant experience of being online. 2022 became a year in which I explored this idea and derived some insights as to how best to manage it.

It must be confessed though that this paper represents a kind of declaration of surrender. Or at least, a form of contentious objection. I don't think we can fix the Internet. I don't think any degree of constructive engagement within our various communities of practice can overcome the deep structural financial and attentional incentives that have wormed their way into every bit that flows through every router. I hold no hope of meaningful improvement with regards to online culture. I think that battle is lost. Instead, I provide this paper to outline a method and a philosophy by which we may, as individuals, draw a range of protective sigils around those parts of the online world in which we still find value.

An Auto-Ethnographic Perspective

This is not a robust study of those that have looked to disconnect from the Internet. There do exist some accounts in the academic literature [e.g. 17, 20], but anecdotal accounts of people who have made the attempt are primarily to be found in the popular literature of magazines and in the video libraries of YouTube. Homewood et al. offer a useful lens on the

process with regards to the 'removal of technologies and documenting their absence as a method' [13], which has parallels to the approach taken here even if the domain of application is not comparable. However, most of the popular accounts follow a pattern I decided early on was sub-optimal - they focus on an approach which requires an ongoing cost in terms of self-denial. Self-control requires a constant exertion of willpower [3, 19] and without diligently changing baseline behaviour it is unsustainable for most people as a permanent lifestyle change [8].

My initial goal with this project wasn't just to 'see if I can do it' and then move on. Rather it was to realign my relationship with being online on a permanent and sustainable basis. My 2019 depth year had the effect of re-calibrating my relationship with capitalism – over the course a year I went from buying things so often that Amazon essentially tried to schedule an intervention (Figure 1) to the point where I simply stopped wanting things. Alongside Project Unplug I also ran Depth Year 2022, reasoning that it represented a natural marriage of goals.

It was that level of mental reprogramming that I wanted to achieve with Project Unplug, and as such the approach taken by the self-denial literature is not wholly appropriate even if it does come with some useful pragmatic advice. My experience is that a healthy relationship to the Internet requires an intense self-evaluation of what one **wants** to get out of the Internet. It requires self-interrogation of what brings happiness as opposed to what simply provides pleasure. It is an approach that begins, first and foremost, situated in the messy context of real life. It is a process that seems like best explored from an autoethnographical perspective.

The act of autoethnography has increased in popularity in the wider academic literature over the past two decades across multiple domains [1, 9, 24]. There is an old piece of received wisdom in scientific publishing - 'the plural of anecdote is not data'. In that respect, an autoethnography's position as inherently anecdotal seems at odds with the mores of the traditional academic model as it is understood in many fields - particularly the fields of qualitative user study with which I am most familiar. The perfect response to this could be argued to come from Sara Pikelet on Twitter, - 'they're not anecdotes. They're small-batch artisanal data'. However, the tension between these

Hi [REDACTED] You're receiving this email because you are purchasing on Amazon in large amounts. Are you purchasing for business?

Figure 1. An email from Amazon, threatening what looks like an intervention.

two divergent perspectives shows a need to validate, at least for the time being, the value of autoethnography in academic papers.

Pikelet's response is a good place to start as it recognises the importance of the small scale, personally constructed narrative. Anecdotes cannot offer broad general insight on larger systemic issues. They are too mired in personal biases, in personal contexts, and in personal mindsets. The second-person techniques that drive much qualitative research have corrective procedures built into them to permit a researcher to draw actionable insight from multiple personal perspectives. Surveys, focus-groups, interviews and so on must be constructed and conducted in ways that eliminate as much bias as possible. We might think of the difference in terms of error bars. In a personal anecdote, the error bars are unknown - both in range and intensity. With carefully designed second-person techniques those error-bars can be made visible and their impact controlled for in analysis. The autoethnography, methodologically, is made up of nothing but invisible error bars. It is also though an impeccable form of engaging with personal meaning [15] and lived experience [22]. Where the autoethnographic technique shines most brightly is in terms of capturing true authenticity [30]. Its focus on self-usage of technology and self-interrogation of experience [7, 12, 25] gives a tool through which otherwise unobtainable perspectives can be codified in the academic literature. The format has its inherent limitations [14, 29] - in order to have the most genuinely authentic research data one must accept inherent constraints on generalisability. It is though in the very tight coupling of the personal to the impersonal that creates the value of autoethnography. The observations associated with that union is the research output, and no part of it can survive deconstruction. That is both its methodological weakness and its methodological value - it is in the words of Pascal 'one of the dominant and characteristic forms of literary self-expression' [23].

3 The Personal Context

To contextualise this within the personal, it is necessary to provide some minor biographical details about my lifelong relationship to the Internet. Before there was such a concept as a digital native, I was already assembling my immigration papers for cyberspace. I was a mental refugee from the physical world long before there was a place for me to go. I got my first computer when I was four years old - a ZX Spectrum that had been handed down to me when my father upgraded to a Commodore 64. I took my first steps into programming when I was five. I wrote my first game (a quiz game which stole liberally from the Trivial Pursuits board game) when I was six or seven. The earliest memory I have is of the game Manic Miner, and being enchanted by the fact that the 'music' kept playing even when the television to which the Spectrum was connected was switched off. If there is one iron rod that has underpinned my personality from childhood to adulthood it's that I don't really enjoy 'going places' or 'doing things'. Before the Internet became a background hum of

modern living, I'd collect up my pocket money and then connect up to Bulletin Board Systems (BBSes) in the evenings when the cost of a phone-call was lowest. When dialup Internet became popular, my habits changed to spending my money on my Internet hours. When the Internet first reached British homes it was usually accessed through a modem and a local telephone number and you had to pay per minute when you were connected to your Internet Service Provider (ISP). Per-minute costs during the day - when the lines were most in use - was often astronomical. During the evening it was much less, and overnight it was lowest. I spent a lot of happy nights browsing through the cornucopia of weirdness that characterized the early World Wide Web.

I should probably also mention here that I am a life-long insomniac. There is probably a causal link, but for the life of me I can't imagine where.

As cost to connect diminished to a mere one-off monthly expense, availability of the Internet increased. So did the time I spent on it. My largest experience as a game developer has been in Multiuser Dungeons (MUDs) [2, 2], and common to that community of developers is a mindset of being 'always connected'. It didn't matter if you were actively developing, you'd be 'idling' and could likely be roused with nothing more than a mention of your name on a chat channel. You were, in other words, always at the computer which usually meant 'always online'. This wasn't a problem. This was joyful. This is where most of my fondest memories can be found - in what to other people would look like indistinguishable mental snapshots of the command line interface of an obsolete operating system. These memories though make me laugh out loud, or tear up, or remember lost friends with fondness. Externally this behaviour may have been troubling, but for me it was the core of a happy existence.

Something though has changed in the tenor of the Internet. Or perhaps more accurately, something has created a disharmony between myself and the online existence I have long cherished. Perhaps this is a case of 'It's not you, it's me'. I don't think so though. I think 'Sorry, no - it's definitely you'.

For quite some time now I have found that social interactions online bring me down rather than cheer me up. And I think in many cases it's simply due to the fact that human civility does not seem to work at Internet scale. The morbidities of online life (the trolling, the spamming, the outright hostility, the polarization of discourse, the alarming resurgence of the far-right, the self-righteous moral policing of the far-left, all of it) are not simply unfortunate. They're built as inevitable consequence into the dynamics of what happens when millions of people can communicate frictionlessly within a shared digital space. MUDs and BBSes had one precious properly encoded into their technical architecture - they could only scale up to hundreds, not millions. That allowed for culture to develop, norms to be established, and rowdy newcomers to be inculcated into a wider context. That inculcation was vital particularly during the September period where undergraduates first gained easy access to the Internet and began the

process of acclimatizing themselves to online culture. As the Internet became generally more available, the psychological September extended over longer and longer stretches of the calendar, until it became perpetual. The Cassandra-like pronouncements of Internet elders - that we were on the verge of a September that Would Never End [16] were not heeded. Now there is no longer an opportunity for newcomers to cyberspace to be educated in a core set of shared values.

This is a somewhat misty-eyed and rose-tinted account of the past. Everything problematic with the Internet now existed then but as has often been remarked, 'million-to-one events happen every minute on the Internet'. If one in a thousand people on a MUD was a crypto-fascist looking for recruits, they had little chance of finding another in a community that was only ever several hundred strong. Their potential allies were still there on the Internet, there were probably as many of them as there are now, but they could never cohere because they were distributed too widely through too many services. Internet scale creates endless opportunities for this coherency and makes them mathematically inevitable. Smaller scale communities are antibodies against cancerous elements multiplying.

And it is this more than anything else that has been making me sad when I sit down to the Internet. I no longer find friendly communities of like-minded souls. I find battlegrounds. I no longer discover forgotten archives of fascinating hobbyist enthusiasm. I find graveyards. The Internet of Beefs¹ is now threaded through every facet of online discourse, and it has wearied me to the point of depression. I termed this the 'outrage economy' and along with surveillance capitalism it is one of the dominant features I can observe driving online interaction. Thus, Project Unplug 2022 - the quest to find the safe harbours and havens that exist online, so that I may once again find joy in my online activities. It was, in essence, a desire to construct my own personally optimal 'Cozy Web' through setting firm limits on what could make its way from network cable to screen.

4 Preparation

The first preparatory steps that would become Project Unplug 2022 actually began early in 2021, as part of the broader exercise of thinking about the logistics of the attempt. I had decided that on the 1st of January, 2022, the project would begin in earnest. From that point I would be following a strict diet of abstinence from Internet activities except within firmly defined exemptions - my own personal 'green zones' of the Internet battlegrounds. The first tentative explorations around this reflected a naivety that would not last. I wanted to block the Internet entirely at the level of the router and only let a small subset of websites in. These would be things like Overleaf, Google Docs, Google Scholar, and so on. Essentially those sites that gave me access to necessary services for my job. On top of this I'd whitelist services that created importance convenience - grocery deliveries, online billing, and so on. This is the core philosophy which became Project Unplug 1.0, which I will term

¹<https://www.ribbonfarm.com/2020/01/16/the-internet-of-beefs/>

the 'Whitelist Approach'.

Early consultation with friends resulted in three particularly important exemptions - Facebook, Slack and Discord. I then looked at the sites that brought me happiness, and evaluated whether I needed to experience them online. Sites like YouTube for example have a number of content-creators I enjoy immensely, but the problem with platforms like this is that they are constantly envious of your attention and designed to keep your eyeballs there. The content creators on these platforms are engaged in a Total War for attention. The subscriptions I maintain on YouTube put perhaps four videos a day - at maximum - into my eyeline and of them perhaps two were something I would consciously choose to watch. And yet, I often found myself in a position of traversing from an interesting feature on the harmonic constructions of the Beatles to a clip in which a guy in a tanktop drives a truck over an ice-cream van. YouTube is indifferent to my usage wishes - everything from autoplay to featured videos conspire to shove more content into my face. This is not unique to YouTube of course - Facebook is constantly doing the same. Instagram, Tik Tok, Twitter, Reddit - they are all designed around the attention economy [4] and if you are not wary you'll find your precious life-seconds frittered away on momentary distraction [21]. Facebook in particular would be a problem for my approach, because it intentionally threads distraction amongst the updates from your friends and it is resistant to attempts to constrain it. Taming Facebook became a regular activity of Project Unplug 2022, given it was one of the sites where I felt the value could, with cultivation, outweigh the cost. My approach for YouTube was to add things in which I was interested into a playlist, and periodically download that playlist to my local machine.

Once I had developed a strategy for all the things I needed, and all the things I intentionally wanted, I turned to the 'known unknown' of what I was actually doing online. For that, I installed a time-tracker on my browser and left it running for nine months during 2021. I then looked at the stats and let my sense of disappointment drive the rest of my decisions. I'd compare the time I spent against the time I thought I spent against the time I wanted to spend. If sites were out of alignment with my desires, they got blocked at the hosts level on my computer. I could obviously unblock them any time I liked, but adding in that barrier of intention would make me think about why I'd blocked them in the first place. They'd be blocked away on January 1st.

Next, I thought about the sites that made me directly unhappy. I took an executive decision here - if I felt worse after checking a site than I did before, then I'd block it right away. I wasn't going to spend nine months knowing a site made me unhappy to justify a desire for chronological purity - as soon as I recognised the effect, I would act on it. The two biggest offenders here were Reddit and Twitter, both platforms with a very sharp 'cheerful to depressed' acceleration curve. On Twitter, it only takes a handful of tweets before I feel myself mentally crashing. That's how long before I find someone's ill-informed 'hot take' or someone's massively retweeted snark related to some transient, insular drama. I'd long been disengaged from Twitter, but I still would keep track of conversations and notifications. Similarly with Reddit - it's the same thing except in longform. Those sites got blocked right away, as did some others. I also blocked all news sites, recognising that anxiety is not activism and the witticism likely misattributed to Mark Twain - 'if one does not read the news, one is uninformed. If one does read the news, then one is

misinformed'.

Finally, I switched off all notifications from all devices, and removed the browser from my phone and tablet. Again, I didn't wait to do that. I also stripped away all the apps for services that hadn't made the cut. I removed all my email access on my tablet, but my phone is a work phone and thus it was still attached to email servers. I would have liked to have formally renounced my mobile phone during this project, but it's just not really feasible to do that in Sweden. Electronic identification is done via a phone, and it's a major part of how you interact with Swedish online services. You can buy public transport tickets without a phone, but it's inconvenient. As my Swedish is still abysmal, I also needed access to Google Translate. But my phone would become an intensely utilitarian device and as a result even when I was unoccupied it spent far more time in my pocket than it did in my hand.

At the end of this process, what I had was something I had never articulated before - a 'mission statement' of what I wanted to get out of the Internet. I summarised what I really valued from being online. This may seem like an obvious list that anyone could put together, but I found myself floundering when it first came to the question 'What do you want to get out of this project?'. Interrogating myself and my friends was an important part of the process of 'finding joy'

- (1) Keeping up with my genuine friends
- (2) Making available the tools needed to work, especially from home
- (3) Having access to the resources that let me contribute to my various meaningful personal projects
- (4) Being entertained by a small, curated list of content creators who I felt respected my time.

I resolved that all of this would only apply to my home computer - my work laptop was only something I ever used on site, and given it is centrally managed by our IT department I would find it difficult to ensure my various software tools and browser plug-ins would work seamlessly alongside the update calendar. Also, my home computer is a Windows machine, my work laptop is a Mac, and there was not a service parity across the two platforms. Also, in all honesty, I was less invested in making sure I was optimally using my employer's time.

I instituted a policy for my email. I would check it twice per workday and never on the weekends. This is a common email management technique and seemed to fit neatly into Project Unplug. This policy lasted perhaps three weeks before I felt the need to abandon it entirely. The lag it introduced into executing administrative tasks was too great to bear.

As to that mission statement, reflecting on my internet usage during 2021 revealed some alarming stats. The first is that I was spending around six hours a day on the Internet. Many of the 305 minutes of online time were spent on sites that I did not feel were meaningfully bringing value to my life. This, I stress, is just on my home computer. Some of those minutes were spent well, on things that would be appropriate during Project Unplug 2022. The rest were often minutes spent on activities that I couldn't even recall or on sites that I didn't even recognise. Many minutes of time spent forming only short-term memories about matters that I could no longer even bring to mind. What the stats showed, for a single year, was evidence of an alarming waste of a lifetime.

When the time to begin was fast approaching, I posted an update to my Facebook friends as a public statement of intention and accountability. Two days

later, the Internet shutters fell. My various protective software runes and computational sigils spun themselves up and began orbiting my online behaviour. From the first minute of the new year, I was within a self-created pentagram drawn in blood promises across my operating system.

5 Project UnPlug

In the end, Project Unplug was actually broken into two separate versions on the basis of insights gleaned. Project Unplug 1.0 conformed to the model outlined above, and lasted six months (January 2022 to June 2022). I reviewed and revised the project in line with results, and Project Unplug 2.0 was instituted in July 2022 and carried through to the end of December of that year.

As part of this process, I wrote diaries on its progress on a monthly basis for the subscribers to my Patron, which is linked to some of my hobby based work [2, 2]. Each month came with a one-thousand or so word summary that reflected my thoughts as the project was being undertaken. Quotes in these subsections are taken from those, typos included.

Common to all approaches was a set of software tools I will outline here. I will provide supporting information ontools specific to each version of the project in their corresponding sections.

- (1) Cold Turkey. This is an access blocker that integrates into your browser to stop access to any websites you have indicated. It can work on a whitelist basis or a blacklist basis. For the first half of the project, I used it as a general blacklist with named exceptions based on the preparation work. This blacklist was lifted during work hours, but was fully in place at weekends and evenings.
- (2) A Time Tracker plugin. I used a different one of these for each half of the project, but they were used to measure time I was spending frivolously - as in, time not spent directly on personal project, life-convenience, mindful recreation or work-related services. These were both stocked with exempted sites, since I don't care how much time I spend searching for academic articles, or on our corporate intranet, or shopping for groceries. At the end of each month (with two exceptions), I'd see how much time I was spending online and adjust my various tools accordingly to correct for identified problems.
- (3) Hosts Blocking. Most operating platforms allow you to circumvent the normal lookup architecture of the Internet by redirecting certain URLs. I did this to sites that I didn't want to accidentally enable - Twitter, Reddit, all the news sites and so on. It's possible to unblock these sites, but the awkwardness of it forces you to think about why.
- (4) Facebook Purity. Honestly, this plugin is just revelatory - it lets you turn Facebook from what it is now into what it used to be. You can use it to banish everything from sponsored posts to adverts. It turned my Facebook timeline from an obstacle course to an immaculate (and brief) update of status updates from friends, which is all I want from it.
- (5) Pocket. This is an offline reader to which you can send any interesting article with a click from a browser. It has become not just a way of ensuring mindful attention but also a treasure trove of references to my own passing indulgences. No longer do I say 'I

read an article once, on a thing I don't remember', because Pocket has the article right there.

- (6) Highlight or Hide Search Results. This is a browser plugin that allows you to curate the output of Google's search engine. It is awkward if you want to avoid Quora, for example, and that is the only place Google Search sends you. This allowed me to align search output to project goals.

Each of these tools has become an inseparable element of my post-Unplug online life. The personal value they generate is inestimable.

5.1 Project Unplug 1.0 - January to June

Only one tool was employed in this specific phase of the project - Youtube-DLG. This is a piece of software that allows you to download a Youtube video from a specific URL. If you pass it a playlist, it'll download everything associated with it. The ethical issues associated with this are outside the scope of this paper.

The obvious problem with the Whitelist Approach became clear early on in Project Unplug 1.0 - there is no such thing as a 'website'. There is a URL you access but the various parts of the site are almost always drawn from dozens of different places. Analytics is on one URL, the login authentication on another. JQuery is stored elsewhere. To allow a site safe passage through a hard blacklist is to manage an endlessly changing network of supporting URLs. The first few weeks of using Cold Turkey required constantly tweaking and curating the whitelist to permit the backend architecture of sites to function. Every month would need some of this fine-tuning.

However, other than this small technical issue my first month of digital abstinence went well. I went from over six hours online a day to two hours, and most of what my time-tracker was capturing was 'endorsed behaviour', as in activity on work and project related sites. As time went by the activity capturing focused on time spent 'outside' of endorsed activities. The stats showed in the first month I spent 6h 41m on Facebook, 4h and 22m on Youtube (presumably browsing for things rather than watching things, since I was using Youtube-DLG). But most of the recorded activity was our CMS at XXXXXX XX XXXXX, Google, my web hosting, my research blog and so on. As these sites incrementally got excluded from tracking, the figures for my online time trended much lower over the year as it more tightly conformed to identified 'wasted' time. I also found out that the time tracker and Cold Turkey didn't work well together, and time spent sitting on a blocked website was still counted as time used. In my first update on the project, I said this:

The 1st of January 2022 fell on a Saturday which means I hit the ground running. From the moment the clock hit 00:00 my internet blocker slammed into effect, with a 56 hour window during which the vast majority of the Internet would be unavailable. And I found that I didn't really miss it much. Don't get me wrong - I kept opening up sites only to be told 'You've quit these' but that was mostly muscle memory. I can't say I leapt into a new philosophy of intense productivity - I was also recovering from what might have been Covid. Mostly I just watched Kurzgesagt videos I had downloaded in 2021. Still though, I expected the start of the process to be bad. I expected to get Digital Withdrawal Symptoms. Not a bit of it

though.

In February, I have begun to refine my methodology for time-tracking in this project. I've whitelisted a number of sites so they don't record usage. XXXXX intranet websites, funding portals, etc - all those things that

are strictly work related. After all with this project I'm not looking to reduce how much I work, but rather the unproductive and idle clicking around that has characterized a lot of my online time. With those changes, the revised stats for January went down to 84 minutes. At this point, the figures became comparable because at the end of each month I'd report on the stats, and I'd manually recalculate based on the use-case of the web access - whether it was something I wanted to control. Table 1 shows the breakdown of time spent each month.

Month	'Frivolous Minutes Per Day'
January 2022	84
February 2022	23
March 2022	13
April 2022	N/A
May 2022	5
June 2022	13

Table 1. Frivolous internet spend in Project Unplug 1.0

In April I was in the UK for a sizable portion of the month, and only had my work laptop with me. Thus, figures weren't captured reliably enough to be a useful comparator. Instead I treated it as a kind of control month, reflecting on how my first three months had affected my reflexive Internet use:

I said in the previous update though I was going to use it as a kind of checksum for the whole endeavour - it would be a chance to observe whether or not I suddenly fell back into the old pattern of Internet overuse. And thankfully the answer was - no. The truth is, I see less and less value in the Internet as time goes by. It's a treasure trove of great stuff, but those treasures are embedded into a vast landscape of utter banality. It has all started to remind me of that great short story by Jorge Luis Borges about the Library of Babel. It's a place that contains an endlessly expanding collection of books, each consisting of random permutations of the letters of the alphabet. Logically, it must contain all human knowledge - past, present and future. And logically it must also contain a perfectly accurate index to its contents. . . somewhere. And the librarians that go searching for it go insane because of the futility of the quest. Sure, there's an accurate index in there - but there are also infinite inaccurate ones, and even they are dwarfed by the endless array of books that are literal gibberish.

Really the Library of Babel is a cruel joke. It tells us that content, no matter how rare and important, has no value without curation. And in turn, that curation has no value without editing. Sure, there's a book in the library that tells you the secrets of the universe. The problem is that it comes with an infinite number of alternative books that tell you a different story about the secrets of the universe. Which one is correct? Don't worry, there's a book in the library that will tell you. . .

Towards the end of this phase of the project though, something became very obvious - I'd actually gotten very close to the ideal of an optimised Internet experience as part of the preparation. The exercise of digital abstinence essentially communicated to me that taking this extreme approach to online activity had a negative impact. It was a process of alchemy in which time spent online became time spent auditing time spent online. In June, my review became:

I think Project Unplug 2022 has reached a natural end-point for phase one. As I've said in previous updates it just feels like I'm spending more time in the bureaucracy than I'm getting back in benefits. My final review of the approach of 'Shut down the Internet entirely except for some whitelists and except for work hours' is - yes, it's doable. It's not even really difficult. But I'm not sure it's worth it. I've given it six months, and the remaining six months of 2022 will be given over to phase two. Phase one - block the Internet, only let a few things in. Phase two - open the Internet, block things when they take up too much of my attention. What the past six months have shown me though is that you absolutely can significantly disconnect from the Internet and you won't really miss it.

And, a reflection on how the auditing was itself creating its own form of Internet anxiety:

May was obviously the point when I was most aggressively policing my Internet time and it was also the point where I think I was getting the least out of the project. I'd dart in and out of websites like a frantic mouse, trying not to accumulate too much on the time tracker. But looking at the time tracker stats in terms of where the minutes went, I never really saw much listed for the month that I felt was a genuine waste of my time. You can certainly live on five minutes of (non-work) Internet a day. It's just not really worth the hassle. Time spent goes down, but visits go way up.

It was clear at this point that blacklisting the unpleasantness offers far more bang for your buck than whitelisting the value. With the use of the various plugins I was employing to curate my online experience on Facebook, the site had become borderline pleasant. I endeavoured then to expand this approach for the remaining year - banishing the content that I didn't enjoy while leaving everything else untouched. Thus began version 2.0 of Project Unplug. I summarised though the takeaway at the end of 1.0 as a series of guidelines that could meaningfully take you 80-90% of the way towards an optimal relationship with the Internet. These were informal guidelines as I put them to my patrons, so some of the personal aspect of the autoethnographical approach shine through here:

- (1) Block all notifications. On your phone, on your tablet, on your computer. You should be the one that decides when it's time to be updated about things, not the app or the website.
- (2) Stop reading the news. Anxiety is not activism, and the news is more interested in your attention than it is in your education. If you need to know, someone will tell you.
- (3) Install Facebook Purity in your browser. It is a magical treasure.
- (4) Install the Unhook Youtube plugin in your

browser. It paves over all the rabbit holes.

- (5) Get yourself a plugin like Web Activity Time Tracker and watch to see where you spend your time online. Look at the sum of what you're doing and ask yourself which of those sites earned the time you spent on them.
- (6) Get yourself an app like Cold Turkey – it lets you set blocks and time allowances for websites and applications. Whenever you see a site leeching time from your life, blocklist it. Or at least ration it – I'm doing that with Facebook now. Two five-minute allowances during the day.
- (7) Alternatively if you've got an old laptop or phone, put all your social media on there and keep it somewhere awkward. Make it effortful to check these things so as to break the harmful patterns of constant refreshing. Strip your phone of all its pointless distractions.
- (8) Consume mindfully. Don't sit on the Internet reading some ridiculous article – save it to Pocket and come back to it when you have the time to really delve into it. Buffer up your Youtube stuff and watch it by appointment. The best place to read most of the articles that the Internet sends your way is on the toilet.

I summed up the takeaways thusly:

Remember you only get this one life, and we're all spending sizable fractions of it online. Budget your life the same way you budget your money. Don't spend it on things that you don't want or need.

5.2 Project Unplug 2.0 - July to December

The second version of the project was more permissive, but still intentionally restrictive in areas where it mattered. I had dealt in Project 1.0 with what I didn't want out of the Internet, and 2.0 then had a focus on what I did want. Facebook, for example, had become much more pleasant through the use of a plugin but I still didn't want to spend hours a day browsing it. It remained consistently my biggest Internet destination in Unplug 1.0, regularly accounting for a quarter of my browsing in a month.

Specific tools employed in this phase of the project were:

- (1) Unhook. This is a browser plugin that removes everything from suggested videos to Youtube comments. It allows you to switch off autoplay, and forces every access to default to your subscriptions page.
- (2) Blocktube. This is a browser plug-in that removes some of the more disruptive content from Youtube. Most valuable for me is that it allows you to block Youtube Shorts, which have turned Youtube into a kind of pound-shop TikTok.

I still used all the general tools, but changed how they were working. Most significantly, I changed my approach to Cold Turkey to reflect the blacklist approach but also to enforce what I thought were fair and reasonable limits on my Internet usage.

The first month in which I employed this approach, there was an unsurprising leap in the amount of time I spent online - it went from 13 minutes in June to 54 minutes in July. However, many of these minutes spent were recorded because of a shift to a different time tracker with richer visualisation options. Re-excluding sites that I saw as being value positive brought that down to 39 minutes per day over the tracked period. It was at this point that I decided thirty

minutes per day was the sweet spot. I felt more at ease saying to people 'I wasted half an hour on the Internet' than I did any larger number. Half an hour in the context of a normal, busy life is virtually nothing. Given the techniques that defined Project Unplug 2022 the average 'value' of each of those minutes was much higher than it had been in previous years. Those minutes have gone from junk-food time to nutritional time. They're apples, not doughnuts.

Also from July 2022:

I am still in a much happier place than I was in 2021 with regards to how much time I spend staring into my nightmare rectangles. If I can hit a sustainable average of thirty 'unproductive' Internet minutes per day by the end of December 2022 I'll consider this project successful. Less would be better, of course, but like any genuine behaviour-breaking process it's important to ensure it doesn't require constant, repeated application of willpower. It should feel effortless.

Month	'Frivolous Minutes Per Day'
July 2022	39
August 2022	19
September 2022	12
October 2022	N/A
November 2022	19
December 2022	19

Table 2. Frivolous internet spend in Project Unplug 2.0

October was a month in which I forgot to reset my time tracker, and by the time I noticed it was too late to gather anything trustworthy in terms of stats. I treated it then as another control month. Aside from the fact I found myself spending more time than I would like on eCommerce sites (which then got their own time management strategy), my notes from the month reflect no other noticeable problems. But more importantly, October 2022 was a chance to reflect on what was likely to happen after Project Unplug concluded. It gave me a chance to think on whether this had been a successful project:

My main realisation from the month has been that there's less about my Internet usage that bothers me than I might have feared.

Given how the project began, this was a remarkably positive outlook. In March of 2021 I was day-dreaming about living a life with no Internet. In October of 2022 I was marked more by a growing sense of comfort - a kind of pleasant indifference that comes only from living safely behind reliable defensive barriers. This attitude progressed into November, when the time tracking showed internet usage comfortably under the 30 minute mark, and December where it was 19. See Table 2 for the full breakdown.

At this point, the project technically ended - but realistically not, because I have resolved to keep my defensive barriers up. I don't feel like I have reset my relationship to the Internet in the same way my Depth Year reset my relationship to consumerism. I feel like if I let myself, I will almost instantly fall back into my lifetime patterns of behaviour. But I do feel that it is possible to construct a safe-space in the Internet with enough self-reflection. I feel less like I am collateral damage in an endlessly futile culture war, and more like the lord of a castle which protects me from the carnage. Its walls may be made of software, but they hold true against the onslaught of the online world.

6 Conclusion

My experience with Project Unplug 2022 has taught me that you don't need to spend six months on a diet of digital abstinence. The value of this project was in

its preparation, and in the blacklist that the preparation identified. Everyone will have their own set of things that make them unhappy, and their own set of things that bring them joy. You can shelter the joyful aspects by blocking out the negative aspects, but it does require an active process of curation. I have made my peace with Facebook by erecting protective runes around its worst excesses. Whether Facebook, Instagram, TikTok, Reddit, Twitter or Snapchat are your platforms of choice, it's important to excise those elements that are vampiric - those that suck more minutes out of your life than you are willing to offer them.

These platforms represent significant relationships in our lives - we often spend as much time with them as we do friends and lovers. The first act of Internet self-care is to set boundaries. The modern Internet economy wants all of your attention, all of the time. Particular platforms want that attention all for themselves and they will employ the digital equivalent of a guilt-trip to make sure your attention minutes contribute to their bottom line. This is understandable in the context of Surveillance capitalism and its associated architecture, but you don't have to submit to it.

I am looking forward to my Internet usage in 2023 being largely optimal. I am perhaps unhappy about 10 or 20% of what's happening to me online, but that is a massive shift from before when I was unhappy about 80% of it. You can put yourself in a position where you are 100% happy with the internet with which you engage, but the cost is extremely high. The approach which I ended with as part of Project Unplug 2.0 gives you a massive chunk of the benefit for what is virtually no cost. I feel psychologically healthier. I feel my habits are sustainable. And I feel comfortable with my relation to the online world for the first time in perhaps a decade.

References

- [1] Tony E Adams and Andrew F Herrmann. 2020. Expanding our autoethnographic future. *Journal of Autoethnography* 1, 1 (2020), 1–8.
- [2] Anon. 2020. Citation Anonymised for Review. *Journal of Anonymity* 1, 1 (2020), 1.
- [3] Roy F Baumeister and Kathleen D Vohs. 2003. Willpower, choice, and self-control. (2003).
- [4] Vikram R Bhargava and Manuel Velasquez. 2021. Ethics of the attention economy: The problem of social media addiction. *Business Ethics Quarterly* 1, 3 (2021), 321–359.
- [5] Li Cheng, Gege Fang, Xiaoxue Zhang, Yuxiang Lv, and Lingxuan Liu. 2022. Impact of social media use on critical thinking ability of university tudents. *Library Hi Tech ahead-of-print* (2022).
- [6] Lisa M Cookingham and Ginny L Ryan. 2015. The impact of social media on the sexual and social wellness of adolescents. *Journal of pediatric and adolescent gynecology* 28, 1 (2015), 2–5.
- [7] Audrey Desjardins and Aubree Ball. 2018. Revealing tensions in autobiographical design in HCI. In *proceedings of the 2018 designing interactive systems conference*. 753–764.
- [8] Angela L Duckworth, Katherine L Milkman, and David Laibson. 2018. Beyond willpower: Strategies for reducing failures of self-control. *Psychological Science in the Public Interest* 19, 3 (2018), 102–129.
- [9] Margot Duncan. 2004. Autoethnography: Critical appreciation of an emerging art. *International journal of qualitative methods* 3, 4 (2004), 28–39.
- [10] Ira Bruce Gaultney, Todd Sherron, and Carrie Boden. 2022. Political polarization, misinformation, and media literacy. *Journal of*

- Media Literacy Education 14, 1 (2022), 59–81.
- [11] Jonathan Haidt and Nick Allen. 2020. Scrutinizing the effects of digital technology on mental health.
- [12] P Holmes and G O'Neill. 2010. Autoethnography and self-reflection: tools for self-assessing intercultural competence. Cambridge Scholars Publishing.
- [13] Sarah Homewood, Amanda Karlsson, and Anna Vallgård. 2020. Removal as a Method: A Fourth Wave HCI Approach to Understanding the Experience of Self-Tracking. In Proceedings of the 2020 ACM Designing Interactive Systems Conference (Eindhoven, Netherlands) (DIS '20). Association for Computing Machinery, New York, NY, USA, 1779–1791. <https://doi.org/10.1145/3357236.3395425>
- [14] Alecia Y Jackson and Lisa A Mazzei. 2008. Experience and “I” in autoethnography: A deconstruction. *International review of qualitative research* 1, 3 (2008), 299–318.
- [15] Stacy Holman Jones, Tony Adams, and Carolyn Ellis. 2016. Introduction: Coming to know autoethnography as more than a method. In *Handbook of autoethnography*. Routledge, 17–48.
- [16] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an “Eternal September” How an Online Community Managed a Surge of Newcomers. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 1152–1156.
- [17] Susann Kohout and Christina Schumann. 2020. Five days without a smartphone. Smartphone use and subjective well-being: Results from a study comparing normal use with a deprivation condition. *SCM Studies in Communication and Media* 9, 3 (2020), 394–420.
- [18] Sabrina Sze Man Lam, Stephen Jivraj, and Shaun Scholes. 2020. Exploring the relationship between internet use and mental health among older adults in England: longitudinal observational study. *Journal of medical Internet research* 22, 7 (2020), e15683.
- [19] George Loewenstein. 2000. Willpower: A decision-theorist’s perspective. *Law and Philosophy* (2000), 51–76.
- [20] Andrés Lucero. 2018. Living without a mobile phone: An autoethnography. In Proceedings of the 2018 Designing Interactive Systems Conference. 765–776.
- [21] Merja Myllylahti. 2018. An attention economy trap? An empirical investigation into four news companies’ Facebook traffic and social media revenue. *Journal of Media Business Studies* 15, 4 (2018), 237–253.
- [22] Faith Wambura Ngunjiri, Kathy-Ann C Hernandez, and Heewon Chang. 2010. Living autoethnography: Connecting life and research. *Journal of research practice* 6, 1 (2010), E1–E1.
- [23] Roy Pascal. 2015. Design and truth in autobiography. Routledge.
- [24] Lynette Pretorius. 2023. A harmony of voices: The value of collaborative autoethnography as collective witnessing during a pandemic. In *Research and Teaching in a Pandemic World: The Challenges of Establishing Academic Identities During Times of Crisis*. Springer, 25–33.
- [25] Rituparna Roy and Shinya Uekusa. 2020. Collaborative autoethnography: “Self-reflection” as a timely alternative research approach during the global pandemic. *Qualitative Research Journal* (2020).
- [26] Ishrat Shahnaz and AR Karim. 2014. The impact of Internet addiction on life satisfaction and life engagement in young adults. *Universal Journal of Psychology* 2, 9 (2014), 273–284.
- [27] Irena Stepanikova, Norman H Nie, and Xiaobin He. 2010. Time on the Internet at home, loneliness, and life satisfaction: Evidence from panel time-diary data. *Computers in Human Behavior* 26, 3 (2010), 329–338.
- [28] Jean M Twenge, Jonathan Haidt, Andrew B Blake, Cooper McAllister, Hannah Lemon, and Astrid Le Roy. 2021. Worldwide increases in adolescent loneliness. *Journal of adolescence* 93 (2021), 257–269.
- [29] Sarah Wall. 2008. Easier said than done: Writing an autoethnography. *International journal of qualitative methods* 7, 1 (2008), 38–53.
- [30] Adam Wiesner. 2020. Contemplating reflexivity as a practice of authenticity in autoethnographic research. *The Qualitative Report* 25, 3 (2020), 662–670.
- [31] Heather Cleland Woods and Holly Scott. 2016. # Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of adolescence* 51 (2016), 41–49.

Michael James Heron
University of Gothenburg
Gothenburg, Sweden
heronm@chalmers.se
pauline.belford@chalmers.se

QUICK-TAKE

Fair Machine Learning Post Affirmative Action

BY NICK PERELLO
PRZEMYSŁAW GRABOWICZ

The U.S. Supreme Court, in a 6-3 decision on June 29, effectively ended the use of race in college admissions [1]. Indeed, national polls found that a plurality of Americans – 42%, according to a poll conducted by the University of Massachusetts [2] – agree that the policy should be discontinued, while 33% support its continued use in admissions decisions. As scholars of fair machine learning, we ponder how the Supreme Court decision shifts points of focus in the field. The most popular fair machine learning methods aim to achieve some form of “impact parity” by diminishing or removing the correlation between decisions and protected attributes, such as race or gender, similarly to the 80% rule of thumb of the Equal Employment Opportunity Commission. Impact parity can be achieved by reversing historical discrimination, which corresponds to affirmative actions, or by diminishing or removing the influence of the attributes correlated with the protected attributes, which is impractical as it severely undermines model accuracy. Besides, impact disparity is not necessarily a bad thing, e.g., African-American patients suffer from a higher rate of chronic illnesses than White patients and, hence, it may be justified to admit them to care programs at a proportionally higher rate [3]. The U.S. burden-shifting framework under Title VII offers solutions alternative to impact parity. To determine employment discrimination, U.S. courts rely on the McDonnell-Douglas burden-shifting framework where the explanations, justifications, and comparisons of

employment practices play a central role. Can similar methods be applied in machine learning?

In machine learning, explanations of model decisions rely on measurements of the influence that each feature has on model outcomes. Such model explanations enable meaningful comparisons between models by highlighting differences in the influence of features, including protected attributes such as race and gender, and the features correlated with them. For instance, take a biased loan approval system where race and zip-code are available and correlated with each other. When race is used for a model’s decision-making, model explanations (such as feature influence measures) will reveal that race is impacting the model and, therefore, the model is discriminatory. If race is removed from the training data and the model is retrained, the explanations will show that the impact of zip-code increases as it is being used as a proxy for race, i.e., the model is “redlining”. To prevent discrimination, fair machine learning and model explanations can be combined to simultaneously drop the impact of the protected attributes, while maintaining the impact of the remaining features [4]. This approach does not rely on impact parity and presents a middle ground between not conducting any discrimination prevention and impact parity methods equivalent to affirmative action. Most importantly, this approach offers a middle ground between the protected, possibly intersectional [5], groups, since it does not disadvantage, nor advantage, any of the groups. Such fair learning methods seem to better align with the existing legislation, the Supreme Court decision, and a democratic decision-making.

References:

- [1] Students for Fair Admissions v. Harvard, 600 U.S. 181, Docket No.20-1199. 2023. Supreme Court of the United States.
- [2] Tatishe M. Nteta, Raymond La Raja, Jesse Rhodes, Alex Theodoridis, Doug Rice, Lane Cuthbert, Adam Eichen, Maddi Hertz, and Donald Snyder. 2023. Juneteenth National Poll. University of Massachusetts Amherst 2023. Retrieved from https://polsci.umass.edu/sites/default/files/CrosstabsNationalJuneteenth2023_0.pdf.
- [3] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453.
- [4] Przemysław A. Grabowicz, Nicholas Perello, and Aarshee Mishra. 2022. Marrying Fairness and Explainability in Supervised Learning. ACM Conference on Fairness, Accountability, and Transparency (FAccT) (2022).
- [5] Przemysław A. Grabowicz, Nicholas Perello, and Kenta Takatsu. 2023. Learning from Discriminatory Training Data. AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) (2023).

NickPerello
Przemyslaw Grabowicz
UMass,

Classical Solutions for Quantum Challenges: An Introduction to Postquantum Cryptography

BY BRAD LONG

Quantum computing has tremendous potential to change the world by solving many previously unsolvable problems. However, with this tremendous computational power comes threats to our existing technologies safeguarding the world's communication channels and data storage. Specifically, it threatens our standardized and widely deployed cryptographic systems that are in use today. These existing cryptosystems are based on mathematical techniques that are difficult (essentially infeasible) for a classical computer to solve. Quantum computing presents a threat since many previously infeasible problems are likely to become feasible or even easy to solve by a quantum computer. This paper briefly reviews quantum computing and its properties before studying related work into postquantum cryptography. The eventuality of quantum computing is discussed based on readily available research and public information, in addition to expert opinion, which then provides insight into the eventuality of postquantum cryptography and the validity of action, or inaction, around this research. Further topics focus on the standardization of postquantum cryptosystems, and future research trends and opportunities in the field of postquantum cryptography.

Keywords: Security, quantum computing, postquantum cryptography, cryptanalysis
 Categories: Security and privacy ~ Cryptography ~ Cryptanalysis and other attacks
 • Computer systems organization ~ Architectures ~ Other architectures ~ Quantum computing

1. Introduction

Research into quantum computing has been increasing at a rapid rate over the last decade and, in particular, the last few years with interesting from commercial organizations such as Microsoft, Google, and Amazon all announcing research into building a quantum computer or quantum chip [18, 24]. Quantum computing's tantalizingly extraordinary computational power means problems that were previously practically infeasible may become easily feasible. This enhanced computing power will provide many benefits across society, but also introduces new threats. Ever since Shor [8, 11] formulated his quantum algorithm for factoring large prime numbers in polynomial time, the cybersecurity community took notice, and has been researching the effect of quantum computing on a range of cryptosystems. At an unspecified future point in time (which may be approaching quicker than expected), quantum computing is expected to break the most widely deployed cryptosystems available today. However, no efficient quantum computer currently exists, and experts provide wide and varied predictions of quantum computing availability. Also worth noting is that alternative quantum-resistant cryptographic techniques exist that may be a secure alternative in a quantum computing world and should continue to be researched, studied, and investigated. In an effort to determine the state of play into postquantum cryptography, this paper begins by providing some background information on quantum computing and quantum cryptography, and then examines potential postquantum techniques, and subsequently looks at existing and future research opportunities in the field. Specifically, Section 2 covers background in quantum computing and quantum properties. In Section 3, a literature review is conducted into alternatives that could potentially provide postquantum secure computing. Section 4.1 discusses quantum computing and its imminence, Section 4.2 introduces post-quantum cryptography, Section 4.3 discusses the standardization of postquantum cryptography and finally, Section 4.4 covers research trends and opportunities.

2. Background

2.1 What is Quantum Computing?

The research into quantum computers has been fast-paced and very productive in recent years [4]. But what is a quantum computer and why does it pose a threat to existing cryptosystems? Whereas a classical computer is based on bits that assume the discrete value of 0 or 1 at a point in time, a quantum computer is based on qubits that can simultaneously assume values of 0 and 1 (and in varying proportions) until evaluated, at which time it will yield precisely 0 or 1, in other words, the qubit "exists in a linear combination of states 0 and 1 and cannot be read (or measured) without collapsing its state" [11]. Following are some properties of quantum computers and their explanations.

2.1.1 Superposition

A classical bit can only be in the states corresponding to 0 or 1, whereas a qubit may be in a superposition of both states. For example, two bits in a classical computer can be in four possible states (00, 01, 10, or 11), but only one of those combinations at any time. In a quantum computer, two qubits can also represent the same four states (00, 01, 10, or 11) but because of superposition, the qubits can represent all four simultaneously. Even if more bits are added to a classical computer, it can still only deal with one state (or combination of bits) at a time, however, as qubits are added, the power of the quantum computer grows exponentially. If a quantum computer has n qubits, 2^n states can be represented simultaneously [19]. To put qubits into superposition, researchers manipulate them using precision lasers or microwave beams. A final calculation emerges only once the qubits are measured, which immediately causes their quantum state to collapse (i.e. evaluate/resolve) to either 0 or 1 [20].

2.1.2 Entanglement

Quantum entanglement is a phenomenon where the states of two or more objects must be described with reference to each other, even though the individual objects may be spatially separated. Researchers can generate pairs of qubits that are entangled, meaning the pair of qubits exist in a single

quantum state. Changing the state of one of the qubits will instantaneously change the state of the other one in a predictable way. This happens even if they are separated by long distances [20]. Imagine a pair of quantum particles that start off with a total of 100 units of energy. One has 40 units of energy and the other must then have 60 units of energy due to the law of conservation of energy. Since they are entangled as a pair, once the energy of one particle is measured, the energy of the other particle is instantaneously known regardless of the location of the particle, which could be anywhere, even on the other side of the galaxy [21].

2.1.3 Measurement and Other Considerations

The quantum system's state is not precisely known before measurement since it exists in a superposition of states. Measurement forces the qubit to collapse to either 0 or 1 [11]. Properties that challenge development of quantum computers include decoherence, that is, the loss of information with the environment, and the no-cloning theorem, that is, a theorem that states it is not possible to perfectly copy an unknown quantum state [11].

The properties of superposition and entanglement enable the incredible power of quantum computers. That tremendous computing power poses threats to existing cryptosystems, both in terms of the sheer power of calculation and new quantum algorithms being developed. This could mean that cryptographic techniques that rely on problems that are hard to calculate and deemed practically infeasible, may become feasible, if not easy to solve.

2.2 Postquantum Cryptography

Before going any further, the definition of postquantum cryptography needs to be defined. In this paper, postquantum cryptography refers to conventional cryptographic algorithms based on mathematical problems other than factoring and discrete logarithms, and that are believed to be secure against quantum attacks [2]. That is, cryptographic techniques using current, classical (i.e. non-quantum) computations that purport to be secure in a quantum world. We will not discuss or review quantum

cryptography such as quantum-key distribution (QKD) [23] or other cryptographic systems based on quantum technologies.

3. Literature Review

As established, development of a quantum computer poses a threat to existing cryptographic techniques for securing communications and data due to the computational power available to a quantum computer. In fact, quantum algorithms have already been developed to show how existing techniques, specifically RSA and Elliptical Curve Cryptography (ECC) can be compromised by a quantum computer with sufficient computing power [2, 11]. Given existing, widely implemented and standardized techniques are vulnerable, there are several approaches that show promise as being resistant to the extraordinary computational power of a quantum computer. There is much research into the cryptanalysis and postquantum resistance of the following four cryptographic techniques: 1) code based, 2) hash based, 3) lattice based, and 4) multivariate.

Code based cryptosystems use error correction codes to generate public keys from private matrices with errors purposely injected [11]. They are appealing since the encryption and decryption is relatively fast. In particular, the McEliece public-key encryption scheme [27] was proposed almost 40 years ago and has stood up to cryptanalytical scrutiny since then [3]. The scheme's security relies on two computational assumptions: 1) generic decoding is hard on average, and 2) the public key (a generator matrix) is hard to distinguish from a random matrix. Hard meaning it cannot be efficiently solved, or in fact, intractable given the sufficiently large keys. Although the suitability of code-based cryptosystems has been hindered by the requirement of relatively large key sizes, the McEliece variant with some adjustments (e.g. hidden Goppa codes) has been recommended by Europe's Post-Quantum Crypto Project because it has been studied since 1978 and no feasible quantum or classical vulnerability has been found [11, 28].

Hash based systems were introduced in 1979 by Merkle [8]. Essentially, hash-based schemes combine many one-time key pairs into a single structure using a hash tree. A hash tree is a hierarchical data structure that repeatedly uses a hash function and concatenation to compute nodes. Hash-based (and code-based) schemes are not known to be vulnerable to quantum computing advancements. However, the suitability of hash-based systems is less than ideal since established systems generally require single-use private keys or limit their use [11].

Lattice based cryptography was introduced almost two decades ago in both mathematics and computer science. Around 1996, Ajtai and Dwork introduced a public-key encryption solution based on the difficulty in calculating lattice problems related to the shortest vector problem [6, 15]. No known algorithm (quantum or classical) exists that can solve this in polynomial time. There are a great number of schemes being proposed and developed based on lattices, and it is a very active research field. To improve efficiency, lattice-based schemes often use ideal lattices, the trade-off to this advantage being the reduced problem space (over a smaller set of all lattices). The ring-TESLA scheme [1] is very promising due to its efficiency and proof [8]. Recent work attempts to incorporate lattice schemes into popular computer languages (such as JavaScript) [22].

Multivariate based schemes are based on solving multivariate polynomials over a finite field. Solving

systems of multivariate polynomial equations is proven to be NP-complete. An NP-complete problem is one that cannot be solved in polynomial time in any known way [5]. Hence, these systems are candidates for post-quantum cryptography. Like lattice-based schemes, there are numerous multivariate schemes. One example is the Rainbow scheme [8, 29], however, the lack of clarity over provable security of multivariate polynomial-based schemes means more cryptanalytical research is required.

4. Discussion

4.1 Quantum Computing –

Science Fact or Science Fiction?

To investigate the eventuality of postquantum cryptography, firstly consider quantum computing. Experts in the field estimate the development of a quantum computer to be from 10-15 years away [2, 10, 11, 25] some estimate between 30-50 years [26] and others estimate that a practical quantum computer will never be developed [11]. Although a universal quantum computer is estimated to be many years away, the first quantum computers are likely to be single purpose, for example, specific to a particular algorithm. This means that quantum computing could be used to specifically implement Shor's algorithm, for example, to compromise existing cryptographic techniques. The amount of funding and research effort being directed into quantum computing research by Microsoft, Google, and Amazon [18] adds a great deal of credibility and validity to quantum computing research and the eventuality of a working quantum computer. With such effort, it seems that quantum computing may come sooner than expected.

4.2 Postquantum Cryptography –

Sooner or Later?

With advances in quantum computing continuing [17] a focus on postquantum cryptography seems justified and necessary. In fact, in 2015, the NSA sent an advisory to organizations to hold off implementing Elliptical Curve Cryptography because quantum-resistant algorithms would replace RSA and ECC implementations [4, 11]. This messaging could be an indication that quantum computing might be achieved sooner than expected.

Keep in mind that original estimates were that a one billion qubit quantum computer would be required to crack RSA security, and that estimation was downsized to the requirement of a 20 million qubit quantum computer [12]. This reduction by a factor of hundreds is likely to be reduced further as quantum computing is more fully understood. Some experts estimate that RSA security could be cracked in one day by a quantum computer [12].

Given the uncertainty of when quantum computing could be fully realized and utilized and the high likelihood that it will, at some point, happen, the price of inaction could be devastating. However, deliberate and positive steps are being taken to standardize postquantum cryptographic services [13].

4.3 Standardization –

A Step In The Right Direction

In 2015, NIST called for the standardization and research into postquantum cryptographic techniques [13]. The standardization process is known to take several years (often from 6-10 years). The call by NIST is forward-thinking and necessary so that the standardization process is complete prior to a mature quantum computer being available. Whilst we have discussed four possible candidates in Section 3 Literature Review of this paper, namely, code-based,

hash-based, multivariate, and lattice techniques for quantum-resistance, the most likely candidate to be adopted globally is the one that can be a drop-in replacement for existing standards [4]. The adoption of postquantum cryptography will be far quicker and simpler across industry around the globe if the infrastructure changes are minimal. That is, ideally any new protocols are not too dissimilar to existing protocols (e.g. for key distribution, handshaking, etc.). With that in mind, the current leading candidate appears to be lattice-based due to the breadth and depth of research into the cryptanalysis of this technique and its practicality, however, there is still a long way to go and much more research required. This is especially the case since not all potential postquantum cryptographic techniques have been proven unbreakable by a quantum computer [11].

4.4 Research Trends and Opportunities

The Section 3 Literature Review outlines research trends around postquantum cryptography with respect to enhanced techniques to combat the advanced computing power that will be available with quantum computers. Further research in cryptanalysis of these postquantum techniques is required [7] to ensure that when quantum computers are available, they stand up to the incredible computing power. This additional research is particularly important since no quantum computer exists yet. That is, ideally these postquantum cryptographic techniques would be provably secure, in which case reliance on testing these techniques on an actual quantum computer becomes unnecessary. Therefore, definite research opportunities exist around provably secure cryptographic algorithms and cryptosystems. Other research opportunities include research into cryptographic algorithms that may yet be discovered. For example, in 2017, Yoo introduced an isogeny-based scheme [16] that relies on the hardness of finding unknown isogenies between pairs of elliptic curves. An efficient quantum algorithm to solve this problem is yet to be discovered [8].

Research into new quantum algorithms may significantly affect the security and effectiveness of cryptographic systems developed with classical computers. For example, the quantum algorithm for the principal ideal problem has applications to lattice-based cryptography. A new method for solving discrete optimization problems on quantum computers called Quantum Approximate Optimization Algorithm (QAOA), and a new method for solving exponentially large systems of linear equations called the Harrow-Hassidim-Lloyd (HHL) algorithm may have an impact on the strength of cryptographic systems developed with classical computers [7]. Although lattice-based systems are standing up to quantum scrutiny, researchers have noted relationships between mathematical methods used in lattice-based cryptography and those used in quantum algorithms. Researchers have been focusing on these connections to exploit lattice-based schemes. Ironically, this led one researcher (Regev) to one of the strongest security guarantees of lattice-based systems [7]. There has also been recent research into quantum superposition attacks. This is where an adversary prepares quantum superposition states and injects them into parts of a cryptosystem that normally handles classical data. Some of these attacks are feasible and some are less so, but it brings into question the fundamentals of classical cryptography in a quantum world [7].

Clearly, further work and research opportunities exist in at least the following four areas: 1) new

cryptosystems and cryptographic protocols, 2) cryptanalysis and provably secure systems, 3) standardization of cryptographic systems, and 4) techniques for incorporating proposed cryptosystems into existing infrastructure.

4.5 A final note on quantum computing and postquantum cryptography

It is possible that quantum computing will never eventuate, however, considering the activity in the field, and the analysis available, and presented in this paper, it is likely quantum computing will become a reality (and sooner than one might expect). The cost of doing nothing is too great, given the likelihood of an eventual quantum reality. It is imperative to continue researching, developing, implementing, and standardizing postquantum cryptographic techniques before a quantum world eventuates. If postquantum cryptography is not achieved or is insufficient to secure communications and data, then the world will need to rapidly employ quantum cryptography (using quantum technology to solve security concerns). However, there will be far more work required to prove it secure and be incorporated into existing infrastructure. Until that happens, existing communications and data will be vulnerable.

5. Conclusion

Quantum computing poses a threat to existing cryptographic techniques that are widely in use today and help to secure communications, data access and storage. This paper provides some background context into quantum computing and postquantum cryptography, the latter being defined as cryptography that can be computed using classical computers

that aim to be quantum resistant (i.e. not vulnerable to quantum computing or quantum technological advances). The current state of several current cryptographic techniques including code-based, hash-based, lattice-based, and multivariate are reviewed. These techniques still hold promise and are being actively investigated for vulnerabilities in a quantum world (i.e. a world where quantum computing is the norm). The likelihood of quantum computers being developed was discussed, along with the likelihood of successfully implementing postquantum cryptography solutions and integrating them into existing infrastructure, which is a necessity for wide and fast adoption of any new techniques. NSA flagging the move to quantum resistant algorithms and NIST calling for standardization is a clear sign that organizations are taking the threat seriously and trying to establish secure standards for new technologies prior to a universal (i.e. general-purpose) quantum computer becoming widely available. Much new research is required in cryptanalysis of proposed cryptosystems to ensure they stand up to the power of quantum algorithms and quantum computing in general. This paper recommends four areas of research: 1) new cryptosystems and cryptographic protocols, 2) cryptanalysis and provably secure systems, 3) standardization of cryptographic systems, and 4) techniques for incorporating proposed cryptosystems into existing infrastructure. The time to invest in postquantum cryptography research and standardization is now, before a quantum computer is developed. There is still time.

6. Acknowledgments

This paper was written whilst undertaking a graduate certificate degree in cyber security at the University of Queensland, Australia, coordinated by Professor Ryan Ko, Chair and Director of UQ Cyber

Security.

References

- [1] S. Akleyk, N. Bindel, J. A. Buchmann, J. Kramer, and G. A. Marson, "An efficient lattice-based signature scheme with provably secure instantiation," in *Progress in Cryptology – AFRICACRYPT 2016*, pp. 44–60, 2016.
- [2] M. Mosca, "Cybersecurity in an Era with Quantum Computers: Will We Be Ready?" in *IEEE Security & Privacy*, vol. 16, no. 05, pp. 38–41, 2018.
- [3] N. Sendrier, "Code-Based Cryptography: State of the Art and Perspectives" in *IEEE Security & Privacy*, vol. 15, no. 04, pp. 44–50, 2017.
- [4] L. Chen, "Cryptography Standards in Quantum Time: New Wine in an Old Wineskin?" in *IEEE Security & Privacy*, vol. 15, no. 04, pp. 51–57, 2017.
- [5] J. Ding and A. Petzoldt, "Current State of Multivariate Cryptography" in *IEEE Security & Privacy*, vol. 15, no. 04, pp. 28–36, 2017.
- [6] K. Lauter, "Postquantum Opportunities: Lattices, Homomorphic Encryption, and Supersingular Isogeny Graphs" in *IEEE Security & Privacy*, vol. 15, no. 04, pp. 22–27, 2017.
- [7] S. Jordan and Y. Liu, "Quantum Cryptanalysis: Shor, Grover, and Beyond" in *IEEE Security & Privacy*, vol. 16, no. 05, pp. 14–21, 2018.
- [8] D. Butin, "Hash-Based Signatures: State of Play" in *IEEE Security & Privacy*, vol. 15, no. 04, pp. 37–43, 2017.
- [9] N. Koblitz and A. Menezes, "A Riddle Wrapped in an Enigma" in *IEEE Security & Privacy*, vol. 14, no. 06, pp. 34–42, 2016.
- [10] J. Mulholland, M. Mosca and J. Braun, "The Day the Cryptography Dies" in *IEEE Security & Privacy*, vol. 15, no. 04, pp. 14–21, 2017.
- [11] L. Mailloux, C. Lewis II, C. Riggs and M. Grimaldi, "Post-Quantum Cryptography: What Advancements in Quantum Computing Mean for IT Professionals" in *IT Professional*, vol. 18, no. 05, pp. 42–47, 2016.
- [12] M. Giles, "Explainer: What is post-quantum cryptography?" *MIT Technology Review*, Jul. 12, 2019.
- [13] L. Chen et al., "NISTIR 8015, Report on Post-Quantum Cryptography" *Computer Security Division, Applied and Computational Mathematics Division, Information Technology Laboratory, National Institute of Standards and Technology*. Apr., 2016.
- [14] C. Lochridge, "Chicken Little's Quantum Sky" *Over the Horizon*, Jul 23, 2019. <https://othjournal.com/2019/07/23/chicken-littles-quantum-sky/> (accessed Apr. 26, 2021)
- [15] M. Ajtai and C. Dwork, "A public-key cryptosystem with worst-case/average-case equivalence," presented at the *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, El Paso, Texas, USA, 1997. [Online]. Available: <https://doi.org/10.1145/258533.258604>
- [16] Y. Yoo, "A post-quantum digital signature scheme based on supersingular isogenies," *UWSpace*, 2017. [Online]. Available: <http://hdl.handle.net/10012/12408>
- [17] J. Carmody, "Sydney student helps solve quantum computing problem with simple modification," *ABC News*, Apr. 13, 2021. [Online]. Available: <https://www.abc.net.au/news/100064328>
- [18] C. Reichert, "Amazon, IBM and Microsoft race to bring global access to quantum computing," *CNet*, Apr. 29, 2020. [Online]. Available: [https://www.cnet.com/news/amazon-ibm-and-microsoft-race-to-bring-global-access-to-](https://www.cnet.com/news/amazon-ibm-and-microsoft-race-to-bring-global-access-to-quantum-computing/)
- [19] O'Connell, "Quantum computing for the qubit curious," *Cosmos*, Jul. 5, 2019. [Online]. Available: <https://cosmosmagazine.com/physics/quantum-computing-for-the-qubit-curious/>
- [20] M. Giles, "Explainer: What is a quantum computer?" *MIT Technology Review*, Jan. 29, 2019.
- [21] A. Vutha, "What is quantum entanglement?" *Cosmos*, May 17, 2019. [Online]. Available: <https://cosmosmagazine.com/physics/what-is-quantum-entanglement/>
- [22] Y. Yuan, J. Xiao, K. Fukushima, S. Kiyomoto, T. Takagi, "Portable Implementation of Postquantum Encryption Schemes and Key Exchange Protocols on JavaScript-Enabled Platforms", *Security and Communication Networks*, vol. 2018, Article ID 9846168, 14 pages, 2018. <https://doi.org/10.1155/2018/9846168>
- [23] M. Sasaki, "Quantum Key Distribution and Its Applications" in *IEEE Security & Privacy*, vol. 16, no. 05, pp. 42–48, 2018.
- [24] E. Gibney, "Hello quantum world! Google publishes landmark quantum supremacy claim," (in En), *Nature, News* vol. 574, no. 7779, pp. 461–462, 2019.
- [25] L. Greenemeier, "How Close Are We—Really—to Building a Quantum Computer?," *Scientific American*, May 30, 2018. [Online]. Available: <https://www.scientificamerican.com/article/how-close-are-we-really-to-building-a-quantum-computer/>
- [26] T. Simonite, "Microsoft's Big Win in Quantum Computing Was an 'Error' After All," *Wired*, Feb. 12, 2021. [Online]. Available: <https://www.wired.com/story/microsoft-win-quantum-computing-error/>
- [27] R. J. McEliece, "A Public-Key Cryptosystem Based on Algebraic Coding Theory," in *Deep Space Network Progress Report*, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, pp. 114–116, 1978.
- [28] D. Bucerzan, V. Dragoi, H. T. Kalachi, "Evolution of the McEliece Public Key Encryption Scheme," *Innovative Security Solutions for Information Technology and Communications*, pp.129–149, 2017.
- [29] Y. Tan, et al., "Building a new secure variant of Rainbow signature scheme," *IET Information Security*, 10(2), pp.53–59, 2016.

Brad Long

The University of Queensland
Brisbane Australia
bradley.long@uq.net.au

Technology One Ltd
Brisbane Australia
brad_long@technologyonecorp.com

Are You Really Competent To Be A Computer Scientist, A Computing Educator, or Even To Use a Computer? The Missing Societal Competency

BY RICHARD BLUMENTHAL

Keywords: AI, Risk, Pause, ChatGPT, Arms Race, Enforcement
 Categories: • Social and professional topics → Computing / technology policy
 • Social and professional topics → Professional Topics → Computing and Business

The position postulated in this Parting Opinion is quite simple: the answer to the question posed in the title is a resounding NO for many people; no matter which of the categories you're in. Hopefully, you're curious to know why I believe this. If you think I'm crazy or that it simply doesn't apply to you, please read on and submit your counter argument to the next issue of *Computers and Society*. My opinion derives tangentially from my recent work on the CS2023 Steering Committee and pondering, *what's next?*, with respect to computers and society.

The CS2023 Steering Committee, which is charged with creating the next revision of the curricular guidelines for undergraduate Computer Science education, is nearing the completion of its mission with a proposed release of the final version of the curricular recommendations in December of this year. Concepts related to our special interest group are primarily captured in the topics and learning outcomes associated with the Society, Ethics, and Profession (SEP) Knowledge Area. I believe it's safe to say that each member of CS2023 committee supports, to varying degrees, strengthening the relevance of SEP in the computer science curriculum. Towards this end, the committee agreed that all Knowledge Areas will include an SEP Knowledge Unit containing specific topics and learning outcomes that focus on SEP issues related to that area. For example, the Algorithmic Foundations (AL) area, which has units focused on: (i) foundational data structures and algorithms, (ii) algorithmic paradigms, (iii) complexity analysis, and (iv) formal computational models, will also include an SEP unit. The way I see it is that every lecture, for example, focused on topics in the previous algorithmic units should also include time dedicated to SEP-related issues. Historically, including SEP in the computer science curriculum has not been supported by everyone. As the inclusion of SEP in the computer science curriculum provides a nice foundation for framing my position in this opinion, let's take a quick look at the evolution of SEP in the computing curriculum.

Historical Perspective That Frames the Problem

With one exception, *Curriculum 68*, the first ACM curricular report on computer science [2]¹, didn't address societal issues related to computing. Based on data from the so-called Pierce Report [10], CS'68

categorizes students "where some computer training would be useful" into three groups:

- *Group 1*: "those in scientific or professional programs having a substantial quantitative content (e.g. mathematics, physics, and engineering)....,
- *Group 2*: fields where an understanding of the fundamentals of computing is steadily becoming more valuable (e.g. business, behavioral~sciences, education, medicine, and library science), and
- *Group 3*: comprises those undergraduates who are majoring in areas which do not necessarily depend on the use of computers (e.g. music, drama, literature, foreign languages, liberal arts, and fine arts)" [6].

Interestingly, it's the final group, Group 3, for which a societal-related comment is made, "There are many persons who maintain that even these students could benefit from a course which would give them an appreciation of this modern technology and its influence on the structure of our society" [2].

With the publication of *Curriculum 78* [3], we begin to see the SEP problem I'm focusing on in this opinion piece. CS'78 recommended eight required courses in the computer science curriculum, which were conveniently associated with the references CS1 – CS2². Ten additional advanced courses were also highly recommended, but it was recognized by the committee that "few departments are expected to have sufficient resources to offer all, or even a large majority, of them" [3]. Of particular interest is the strongly recommended elective course CS9 *Computers and Society*. The commentary associated with this course begins with "much debate surrounds the role of this course in the curriculum" [3]³. The aspect of the debate that I'm addressing in this opinion piece focuses on the issue "it has been argued that such a course is not a computer science course, but rather should be in the area of social science" [3]. Before I comment on this aspect of the debate, it's worthwhile to briefly examine *Computing Curricula 1991* and the subsequent *ImpactCS Project* both of which added proverbial "fuel to the fire".

Computing Curricula 1991 changed the approach used to specify the curriculum from a course-based to a subject area-based approach [11]. Unlike CS'78 where

the CS9 *Society and Computers* course was highly recommended, but still elective, the newly introduced SEP subject area in CS'91 was recommended as being required. This SEP requirement essentially led to the creation of the *ImpactCS Project* [6, 7, 8], which focused on "defining the core content and pedagogical objectives for integrating social impact and ethics into the computer science curriculum" [6]. Although there was "overwhelming" positive support for the ImpactCS Project, the authors noted an interesting exception, which can be summarized as "[the] most glaring problem is that the proposed subject matter is not computer science..." [7]. Similar to the commentary in the CS'78 report, an argument is made that SEP is not computer science. The *ImpactCS Project* authors, as well as others, both at that time and since, have presented arguments as to why SEP issues are part of computer science. Instead, I'd like to take a slightly different approach, which leads to my purposely strong position in this Parting Opinion.

Requiring SEP Competency

Although I support the position that SEP is part of computer science, debating this issue distracts from the intent of including SEP in the curriculum. Here's my position, even if SEP isn't computer science, it's certainly includes what every computer scientist, or computing professional, needs to know. Specifically, every computing professional, educator, and computer user should be able to assess how their involvement with a computer or computing impacts society. If they cannot, then they are not fully competent at what they are doing. Furthermore, if one is not fully competent, why should society allow them to proceed?

Every computing professional, educator, and computer user should be able to assess how their involvement with a computer or computing impacts society. If they cannot, then they are not fully competent at what they are doing. Furthermore, if one is not fully competent, why should society allow them to proceed?

¹There was an earlier ACM report with preliminary curricular recommendations in 1965 [1] and an educational issue of *Communications of the ACM* in 1964 [4].

²As an aside, this is where the introductory CS1 and CS2 Programming I and II course sequence "abbreviations" originated.

³For both completeness and fairness to the original authors, the complete commentary for the CS9 course is given in Appendix A of this Parting Opinion.

⁴Again, for completeness and fairness to the original authors, their complete counter-argument is given in Appendix B of this Parting Opinion.

We might ask, is computing too dangerous to leave to individuals who are not adequately competent in assessing SEP issues?

Yes, an intentionally strong position based on a variety of assumptions that result in various implications. Though, we might ask, is computing too dangerous to leave to individuals who are not adequately competent in assessing SEP issues? Consider, for example, that, to the best of my knowledge, every country in the world requires obtaining a driver's license in order to operate vehicles on public roads. While there is currently very little evidence to suggest that computer usage has killed more people than automobile usage, there is growing evidence the computer usage has harmed more people and cost society more monetary damage. In the United States, for example, requiring a driver's license followed quickly on the heels of generally available motor vehicles, but in others it took over fifty years to recognize the dangers of operating a motor vehicle [9]. Using this analogy, my position is a little late. Though, licensing professionals in computing isn't new. So, my position isn't as Draconian as it might first appear, but I am asking to what degree might we desire to enforce it?

What might SEP competency look like:

- *Research Publications* ACM could require all research publications submitted for review to require the author(s) to address the SEP issues. If they do not, the article should be rejected since the authors, in my opinion, do not have the competency to be performing research in this computing area. Aside, this appears to be the policy that SIGCSE Birds of a Feather submissions are currently supporting, with respect to diversity and inclusion. If you don't competently address DEI in the submission, the BoF will not be accepted.
- *Computing Educators* Every computing educator should be able to address the SEP areas of their respective computing curricula, as recommended in the curricular guidelines from ACM, IEEE-CS, AAAI, ...
- *Computing Practitioners* Computing practitioners should be licensed (like other professionals, such as Physicians and Lawyers).
- *Computing Users* Purchasing a computer should require some type of licensure/training.

Admittedly, to what degree the above groups should be able to assess SEP issues is an open issue. However, if the computing community doesn't get a handle on what this means, I'm afraid it will be legislated by others, who might have a position even more extreme than the one I'm presenting. I believe we have to land somewhere between the knowledge associated with a graduate degree in philosophy and the catch-phrases suggest by the SEP is not computer science counter-argument appearing in [5].

However, once we begin to settle on what knowledge is necessary, then I don't believe my position is too Draconian. That is, if you cannot assess the impact your research has on society, no matter how great a computing researcher you are, you are simply not fully competent to conduct research in the area. Likewise for the other groups. As to computer users, who are not professionals, I'm a bit mixed. I do believe that a user should not be able to accidentally empty their bank account, which is an impact on society in the sense that the individual is a

member of society. Perhaps computing professionals can help. Again, most people can drive a car as easily as they can use a computer. However, we still require a license to drive, but any child is allowed to use a computer. Sure parents can manage access, which may be supported by the technology itself, but who trusts the parents? Society is divided on the answer to this question with respect to many other decisions. Is unrestricted computer access a fundamental right and perhaps more importantly to what degree is this right supported or limited? I believe these are issues that our SIG should contribute to.

Thus, and as always, I conclude by asking the *Computers and Society* community, what do you think? At the moment, maybe I'm playing the "Devil's Advocate" since I'm swinging wildly on my position. I invite the community to continue this discussion.

References

- [1] Atchison, W. F., Conte, S. D., Hambley, J. W., Keenan, R. A., Kehl, W. B., Navarro, S. O., Rheinboldt, W. C., Schweppe, E. J., & Young, D. M. (1965). An undergraduate program in computer science: Preliminary Recommendations. *Communications of the ACM* 8(9), 543-552.
- [2] Atchison, W. F., Conte, S. D., Hambley, J. W., Hull, T. E., Keenan, R. A., Kehl, W. B., McKluskey, Navarro, S. O., Rheinboldt, W. C., Schweppe, Vivant, W., E. J., & Young, D. M. (1968). Curriculum 68: Recommendations for academic programs in computer science. *Communications of the ACM* 11(3), 151-197.
- [3] Austing, R. H., Barnes, B. H., Bonnette, D. T., Engel, G. L. & Stokes, G. (1978). Curriculum 78: Recommendations for the undergraduate program in computer science. *Communications of the ACM* 22(3), 147-166.
- [4] Cotlieb, C. C. (1964). Educational issue. *Communications of the ACM* 7(4), 205-227.
- [5] Davis, R. L. & Webster, R. W. (1997). Questioning proposed CS knowledge units. *Communications of the ACM* 40(4), 20-21.
- [6] Huff, C. & Martin, D., (1995). Computing consequences: A framework for teaching ethical computing. *Communications of the ACM* 38(12), 75-84.
- [7] Martin, C. D., Huff, C., Gotterbarn, D., & Miller, K. (1996). Implementing a tenth strand in the CS Curriculum. *Communications of the ACM* 39(12), 75-84.
- [8] Martin, C. D. & Wertz, E. Y. (1999). From awareness to action: Integrating ethics and social responsibility into the computer science curriculum. *Communications of the ACM* 29(2), 6-14.
- [9] Nix E., (2016). When was the first U.S. driver's license issued? *History*. <https://history.com/news/when-was-the-first-u-s-drivers-license-issued>.
- [10] President's Science Advisory Committee. Computers in higher education. The White House, Washington, D. C., Feb. 1967
- [11] Turner, J. (1991). Computing Curricula 1991: A summary ACM/IEEE-CS Joint Curriculum Task Force report. *Communications of the ACM* 34(6), 69-83.

Appendix A CS 9 Computers and Society

Here is the complete description of the CS 9 course from Curriculum 78:

"Prerequisite: elementary core material

- (a) to present concepts of social value and valuations;
- (b) to introduce models which describe the impact of computers on society;

- (c) to provide a framework for professional activity that involves explicit consideration of and decisions concerning social impact;
- (d) to present tools and techniques which are applicable to problems posed by the social impact of computers.

Much debate surrounds the role of this course in the curriculum. While few will disagree that professional computer scientists should be instructed to evaluate social issues regarding that which they do, it has been argued that such a course is not a computer science course, but rather should be in the area of the social sciences. Another argument is presented which states that this material is so important that it should not merely be covered in a single course, but instead should be integrated throughout the curriculum. Although this latter argument has validity, it is difficult to insure sufficient coverage of topics when they are scattered throughout a number of courses. As a result it is recommended that this course be considered at least as a strongly recommended elective. If, in fact, the material to meet the above objectives is not covered in the other intermediate and advanced level courses in this program, then this course should be required.

A computer science major taking an advanced level computers and society course would be expected to be familiar with the elementary material described in the previous section. All of that material, however, is not necessarily prerequisite for such a course. The prerequisite should, in fact, be chosen in such a manner that nonmajors would also be able to take the course. A mixture of majors in such a course would provide broadening interchange and would benefit both the computer science students and the other majors. The course should be taught by the computer science faculty, but team-teaching with faculty from other disciplines should be encouraged. The course could be general and treat a number of computer impact topics, or specific, and treat in depth one of the topics (such as legal issues in computing). This recommendation is conditioned on the assumption that instructors who present material on societal impact, whether as an entire course or as part of other courses, will try to include both sides of or approaches to issues without instilling their own philosophical leanings on complex societal issues. For example, certain topics contain political overtones which should be discussed, but which, if not done carefully, can give the material a political science flavor it does not deserve.

A strict outline is not given. The number of topics and extent of coverage as well as the instructional techniques used can vary considerably and still meet the objectives of the course. A term project involving computer applications that are manifested in the local community is strongly recommended. Possible topics, but certainly not an exhaustive list, that could be included in such a course are as follows:

- A. History of computing and technology
- B. The place of the computer in modern society
- C. The computer and the individual
- D. Survey of computer applications
- E. Legal issues

- F. Computers in decision-making processes
- G. The computer scientist as a professional
- H. Futurists' views of computing
- I. Public perception of computers and computer scientists" [3].

Appendix B Questioning Proposed CS Knowledge Units

Here is the complete text of the original authors submission to the CACM editorial forum:

The most glaring problem with the proposal, 'Implementing a Tenth Strand in the CS Curriculum' [5], is that the proposed subject matter is not computer science. The content of the "strand" has no algorithms, no data structures, no mathematical analysis, no computer architecture, neither software development nor hardware design, no computer science theory. In short, the content is devoid of every standard element present in computer science research and education. We invite the reader to carefully examine the proposed knowledge units. The authors cite complexity, consistency, and completeness among recurring concepts but apparently not the complexity of Karp or the consistency and completeness of Godel. The proposed knowledge units properly belong within ethics or philosophy but not computer science. But why stop at ethics? Proponents could argue for additional strands and courses to address the psychology, sociology, and economics of computer science. Every humanistic discipline could be a strand whose title would include ". . . of computer science," requiring the incorporation of additional strands to Computing Curricula 1991. Proposed laboratories include discussing the winners and losers when the clock was invented, reading or viewing a work of science fiction, discussing a computer professional who "has to make a difficult choice," prioritizing a list of organ-transplant candidates, role-playing, and debating. Decide for yourself whether these are valid computer science laboratory experiences or properly belong within other disciplines. It's hard to imagine a computer scientist teaching these things. A course in social and ethical impact of computing may be desirable, but let us ask the philosophy, sociology, and public policy departments to teach such courses. Ethics should be taught by faculty with experience, research interests, and doctoral degrees in ethics, not by computer science professors pretending to be ethicists. It is naive to expect a computer scientist without a graduate degree in philosophy to speak authoritatively about Bentham's Utilitarianism or Kant's Moral Imperative, unless it is to reduce these theories to catchphrases such as "greatest good for the greatest number." Ethical and social concerns may be important, but as debating the morality of nuclear weapons is not doing physics, discussing the social and ethical impact of computing is not doing computer science" [5].

Richard Blumenthal
 Professor and Chair
 Computer and Cyber Sciences Department
 Regis University
 Denver, CO
 rblument@regis.edu

QUICK-TAKE

"All Watched Over By Machines," or AI Ethics, Surveillance, and Pluralism

BY LEVI PULFORD

Right now, multiple conversations are occurring across places as disparate as damp IT basements, lofty executive suites, virtual team meetings, pristine computer labs, and hazy dorm rooms about what AI really is in this present moment and what AI can really become in the near and distant futures. Doomers claim artificial general intelligence is on the horizon, bringing with it widespread job automation, economic disruption, and the collapse of modern-day societies. Utopians posit that AGI will bring unprecedented freedoms; increase access, safety, and democracy; and reshape our relationships among ourselves and with the world into a perfect, self-sustaining harmony. Halfway along this ideological spectrum, between doomer and utopian, lies the pragmatist, who weighs both the benefits and the risks associated with AI as it is deployed today, with an eye toward how AI might also be deployed tomorrow. Whatever super-intelligent AI might be invented five or fifty years from now, it will likely come about through a series of deceptively small and seemingly insignificant steps, as, in the words of Emily Dickinson, Forever is composed of Nows.

Various officials, experts, and policymakers have acknowledged that AI already has the potential to disrupt our lives in myriad ways. In a recent Sunday Times interview, IBM CEO Arvind Krishna expressed the need for people to develop critical thinking skills if they wish to future-proof their careers against AI [1]. It is my opinion that we have found ourselves in an all-hands-on-deck situation, and we need doomers, utopians, and pragmatists, as well as other perspectives existing within and beyond this spectrum, if we are to successfully crack what Nick Bostrom in 2014 referred to as the "control problem," or "the problem of how to control what the superintelligence would do" [2]. Bostrom was thinking about issues of value-loading nearly three years before the formation of AI ethics, a field centered around determining what our values are, how we arrive at them, and how we might apply them toward AI technologies. However, pragmatists such as Rodrigo Ochigame have since argued that AI ethics as a discourse has become co-opted by tech industry leaders seeking to avoid government regulation [3].

In response to a widespread number of cases and controversies, a few U.S. cities and states have begun regulating and outright banning specific AI technologies, such as face recognition, but there is still much to be done on this front. For instance, a recently proposed Massachusetts bill would limit the governmental use of face recognition technology to a centralized office specifically trained to handle the material, ideally mitigating racial bias and inaccuracy [4]. Even though corporate, academic, and military interests have converged on AI ethics in crucial ways, people continue to be exploited by the move-fast-and-break-things ethos dominant across the private sector. This is perhaps made most salient by the experiences of Kenyan content moderators hired by OpenAI [5].

To think critically, as Krishna calls for, requires us to attend to power. According to Foucault, modern society has subjected itself to an extended Panopticon; our need for discipline has transcended the limits of architecture and technology and

transformed how surveillance operates. For Foucault, "visibility is a trap" which automates the labor of surveillance within the disciplinary society [6]. With the rise of personal smart devices and personalized social media feeds, we have traded the Panopticon for a labyrinth of echo chambers. For the poet Richard Brautigan, however, visibility is how we can come to live in a "cybernetic ecology... all watched over by machines of loving grace" [7]. In other words, visibility is not inherently a trap; it also contains opportunity for liberatory recognition. Just as discipline is part of human nature, so too is loving grace.

To navigate our way through the increasingly disorienting labyrinth that is society in the age of AI, it is not so much about where one aligns oneself on the AI ideological spectrum but more importantly how we recognize, voice, and validate our stances while embracing the stances of others. This is, in effect, a call for pluralism in AI ethics. Before we can act as pluralists, however, we must reconcile what Karl Popper first identified in 1945 as the paradox of tolerance. Put simply: a tolerant society must not tolerate intolerance, and yet too often we cannot agree on what constitutes intolerance in human behavior, let alone intolerance in machines. For now, we can only strive to build artificial intelligence with all the loving grace we, ourselves, possess.

References

- [1] Thompson, Polly. 2023. "This is how to futureproof your career in an AI world, says IBM's CEO." Insider. Retrieved from <https://www.businessinsider.com/futureproof-career-ai-world-ibm-ceo-jobs-employment-2023-10>
- [2] Bostrom, Nick. 2014. Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.
- [3] Ochigame, Rodrigo. 2019. "The Invention of Ethical AI." The Intercept. Retrieved from <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>
- [4] Borondy, Kinga. 2023. "How often should police use facial recognition technology? Mass. considering strict limits." Worcester Telegram & Gazette. Retrieved from <https://www.telegram.com/story/news/state/2023/10/13/massachusetts-police-facial-recognition-limits-proposed/71159918007/>
- [5] Rowe, Niamh. 2023. "It's destroyed me completely: Kenyan moderators decry toll of training of AI models." The Guardian. Retrieved October 24, 2023 from <https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai>
- [6] Foucault, Michel. 2008. "'Panopticism' from Discipline & Punish: The Birth of the Prison." Indiana University Press. Retrieved from <https://muse.jhu.edu/article/252435/pdf>
- [7] Brautigan, Richard. 1967. "All Watched Over by Machines of Loving Grace." The Communication Company. Retrieved from <https://www.readingdesign.org/all-watched-over-by-machines>