AI Matters

Annotated Table of Contents

Automation Moderation: Finding Symbiosis with Anti-Human Technology

Jack Bandy

Full article: http://doi.acm.org/10.1145/3175502.3175514

ACM SIGAI Student Essay Contest Winner

AI Conference Reports

Michael Rovatsos

Full article: http://doi.acm.org/10.1145/3175502.3175515

Event reports on recent AI conferences

Links

SIGAI website: http://sigai.acm.org/
Newsletter: http://sigai.acm.org/aimatters/
Blog: http://sigai.acm.org/ai-matters/
Twitter: http://twitter.com/acm_sigai/
Edition DOI: 10.1145/3175502

Join SIGAI

Students $11, others $25. For details, see http://sigai.acm.org/Benefits: regular, student Also consider joining ACM. Our mailing list is open to all.

Notice to Contributing Authors to SIG Newsletters

By submitting your article 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 in the electronic version of this publication
- to include the article in the ACM Digital Library and in any Digital Library related services
- to allow users to make a personal copy of the article for noncommercial, educational or research purposes

However, as a contributing author, you retain copyright to your article and ACM will refer requests for republication directly to you.

Submit to AI Matters!

We're accepting articles and announcements now for future issues. Details are available at http://sigai.acm.org/aimatters.

AI Matters Editorial Board

Contact us: aimatters@sigai.acm.org

Contents Legend

 Book Announcement

 Ph.D. Dissertation Briefing

 AI Education

 Event Report

 Hot Topics

 Humor

 AI Impact

 AI News

 Opinion

 Paper Précis

 Spotlight

 Video or Image

Details at http://sigai.acm.org/aimatters

Welcome to AI Matters, Volume 3, Issue 4

Eric Eaton, Co-Editor (University of Pennsylvania; aimatters@sigai.acm.org)
Amy McGovern, Co-Editor (University of Oklahoma; aimatters@sigai.acm.org)
DOI: 10.1145/3175502.3175503

Welcome to the final issue in our third year!

This issue features the second installment of *essays on the "Responsible Use of AI Technologies"* from the winners of the ACM SIGAI-sponsored student essay contest. In addition to having their essay appear in *AI Matters*, the contest winners received either monetary prizes or one-on-one Skype sessions with leading AI researchers.

This issue's **AI Spotlight Interview** focuses on *Ayanna Howard*, a professor from Georgia Tech, and renowned educator and roboticist. She even founded a startup called Zyrobotics that focuses on educational robotics, combining these two interests.

Life got you down with too many conflicting demands on your time? Sounds like you need to learn about *distributed constraint optimization* through **AI Buzzwords Explained**.

And if that's not interesting enough, go read about the impact of AI on other fields through *using intelligent workflows to analyze artistic style* by Buell et al., or the latest on **AI Policy** or **AI Conference Reports**.

It sure is a bright time for AI. Never saw the sun shining so bright. There are even a collection of *"blue sky" ideas on education* from the EAAI 2017 New and Future AI Educator awardees. Blue skies, smiling at me....

Phew! That's a lot. Sounds like you need to plan out how you're going to read this issue, borrowing some techniques on *adaptive planning* from our **AI Education** column.

Thanks for reading! If you've enjoyed this corny editorial introduction, we appreciate donations in the form of ideas and future submissions to *AI Matters*!

Eric Eaton is a Co-Editor of AI Matters. He is a faculty member at the University of Pennsylvania in the Department of Computer and Information Science, and in the General Robotics, Automation, Sensing, and Perception (GRASP) lab. His research is in machine learning and AI, with applications to robotics, sustainability, and medicine.

Amy McGovern is a Co-Editor of AI Matters. She is an Associate Professor of computer science at the University of Oklahoma and an adjunct associate professor of meteorology. She directs the Interaction, Discovery, Exploration and Adaptation (IDEA) lab. Her research focuses on machine learning and data mining with applications to high-impact weather.

Submit to AI Matters!

We're accepting articles and announcements for future issues. Details on the submission process are available at http://sigai.acm.org/aimatters.

AI Events

Michael Rovatsos (University of Edinburgh; mrovatso@inf.ed.ac.uk)
DOI: 10.1145/3175502.3175504

This section features information about upcoming events relevant to the readers of AI Matters, including those supported by SIGAI. We would love to hear from you if you are are organizing an event and would be interested in cooperating with SIGAI, or if you have announcements relevant to SIGAI. For more information about conference support visit sigai.acm.org/activities/requesting_sponsorship.html.

The 13th Annual ACM/IEEE International Conference on Human Robot Interaction (HRI 2018)

Chicago, IL, USA, March 5–8, 2018
humanrobotinteraction.org/2018/
HRI 2018 is the 13th annual conference for basic and applied human-robot interaction research. Researchers from across the world attend and submit their best work to HRI to exchange ideas about the latest theories, technology, data, and videos furthering the state-of-the-art in human-robot interaction. Each year, the HRI Conference highlights a particular area through a theme. The theme of HRI 2018 is Robots for Social Good. The HRI conference is a highly selective annual international conference that aims to showcase the very best interdisciplinary and multidisciplinary research in human-robot interaction with roots in and broad participation from communities that include but not limited to robotics, human-computer interaction, human factors, artificial intelligence, engineering, and social and behavioral sciences.

The 12th ACM Conference on Recommender Systems (RecSys 2018)

Vancouver, Canada, October 2–7, 2018
recsys.acm.org/recsys18/
RecSys is the premier international forum for the presentation of new research results, systems and techniques in the broad field of recommender systems. Recommendation is a particular form of information filtering, that exploits past behaviors and user similarities to generate a list of information items that is personally tailored to an end-users preferences. RecSys 2018, the twelfth conference in this series, will be held in Vancouver, Canada. It will bring together researchers and practitioners from academia and industry to present their latest results and identify new trends and challenges in providing recommendation components in a range of innovative application contexts. In addition to the main technical track, RecSys 2018 program will feature keynote and invited talks, tutorials, a workshop program, an industrial track and a doctoral symposium.
Submission deadline: May 7, 2018

The 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018)

January 19–21, Funchal, Madeira, Portugal
www.biostec.org
The purpose of BIOSTEC is to bring together researchers and practitioners, including engineers, biologists, health professionals and informatics/computer scientists, interested in both theoretical advances and applications of information systems, artificial intelligence, signal processing, electronics and other engineering tools in knowledge areas related to biology and medicine. BIOSTEC is composed of five co-located conferences, each specialized in a different knowledge area.

Michael Rovatsos is the Conference Coordination Officer for ACM SIGAI, and a faculty member of the School of Informatics at the University of Edinburgh, UK. His research in in multiagent systems and human-friendly AI. Contact him at mrovatso@inf.ed.ac.uk.

An Interview with Ayanna Howard

Amy McGovern (University of Oklahoma; amcgovern@ou.edu)
Eric Eaton (University of Pennsylvania; eeaton@cis.upenn.edu)
DOI: 10.1145/3175502.3175505

Abstract

This column is the fifth in our series profiling senior AI researchers. This month focuses on Ayanna Howard.

Introduction

Our fifth profile for the interview series is Ayanna Howard, Professor and Linda J. and Mark C. Smith Endowed Chair in the School of Electrical and Computer Engineering at the Georgia Institute of Technology.

Bio

Figure 1: Ayanna Howard

Ayanna Howard, Ph.D. is Professor and Linda J. and Mark C. Smith Endowed Chair in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. As an educator, researcher, and innovator, Dr. Howard's career focus is on intelligent technologies that must adapt to and function within a human-centered world. Her work, which encompasses advancements in artificial intelligence (AI), assistive technologies, and robotics, has resulted in over 200 peer-reviewed publications in a number of projects - from assistive robots in the home to AI-powered STEM apps for children with diverse learning needs. She has over 20 years of R&D experience covering a number of projects that have been supported by various agencies including: National Science Foundation, NewSchools Venture Fund, Procter and Gamble, NASA, and the Grammy Foundation. Dr. Howard received her B.S. in Engineering from Brown University, her M.S.E.E. from the University of Southern California, her M.B.A. from the Drucker Graduate School of Management, and her Ph.D. in Electrical Engineering from the University of Southern California. To date, her unique accomplishments have been highlighted through a number of awards and articles, including highlights in USA Today, Upscale, and TIME Magazine, as well as being named a MIT Technology Review top young innovator and recognized as one of the 23 most powerful women engineers in the world by Business Insider. In 2013, she also founded Zyrobotics, which is currently licensing technology derived from her research and has released their first suite of STEM educational products to engage children of all abilities. From 1993-2005, Dr. Howard was at NASA's Jet Propulsion Laboratory. She has also served as the Associate Director of Research for the Georgia Tech Institute for Robotics and Intelligent Machines and as Chair of the multidisciplinary Robotics Ph.D. program at Georgia Tech.

Getting to know Ayanna Howard

How did you become interested in Computer Science and AI?

I first became interested in robotics as a young, impressionable, middle school girl. My motivation was the television series called The Bionic Women — my goal in life, at that time, was to gain the skills necessary to build the bionic women. I figured that I had to acquire combined skill sets in engineering and computer science in order to accomplish that goal. With respect to AI, I became interested in AI after my junior year in college, when I was required to design my first neural network during my third NASA summer internship in 1992. I quickly saw that, if I could combine the power of AI with Robotics, I could enable the ambitious dreams of my youth.

What was your most difficult professional decision and why?

The most difficult professional decision I had to make, in the past, was to leave NASA and pursue robotics research as an academic. The primary place I'd worked at from 1990 until 2005 was at NASA. I'd grown over those 15 years in my technical job positions from summer intern to computer scientist (after college graduation) to information systems engineer, robotics researcher, and then senior robotics researcher. And then, I was faced with the realization that, in order to push my ambitious goals in robotics, I needed more freedom to pursue robotics applications outside of space exploration. The difficulty was, I still enjoyed the space robotics research efforts I was leading at NASA, but I also felt a need to expand beyond my intellectual comfort zone.

What professional achievement are you most proud of?

The professional achievement I am proudest of is founding of a startup company, Zyrobotics, which has commercialized educational products based on technology licensed from my lab at Georgia Tech. I'm most proud of this achievement because it allowed me to combine all of the hard-knock lessons I've learned in designing artificial intelligence algorithms, adaptive user interfaces, and human-robot interaction schemes with a real-world application that has large societal impact — that of engaging children of diverse abilities in STEM education, including coding.

What do you wish you had known as a Ph.D. student or early researcher?

As a Ph.D. student, I wish I had known that finding a social support group is just as important to your academic growth as finding an academic/research home. I consider myself a fairly stubborn person — I consider words of discouragement a challenge to prove others wrong. But psychological death by a thousand cuts (i.e., words of negativism) is a reality for many early researchers. A social support group helps to balance the negativism that others, sometimes unconsciously, subject others too.

What would you have chosen as your career if you hadn't gone into CS?

If I hadn't gone into the field of Robotics/AI, I would have chosen a career as a forensic scientist. I've always loved puzzles and in forensic science, as a career, I would have focused on solving life puzzles based on the physical evidence. The data doesn't lie (although, as we know, you can bias the data so it seems to).

What is a "typical" day like for you?

Although I have no "typical" day, I can categorize my activities into five main buckets, in no priority order: 1) human-human interactions, 2) experiments and deployments, 3) writing (including emails), 4) life balance activities, and 5) thinking/research activities. Human-human interactions involve everything from meeting with my students to talking with special education teachers to one-on-one observations in the pediatric clinic. Experiments and deployments involve everything from running a participant study to evaluating the statistics associated with a study hypothesis. Writing involves reviewing my students' publication drafts, writing proposals, and, of course, addressing email action items. Life-balance activities include achieving my daily exercise goals as well as ensuring I don't miss any important family events. Finally thinking/research activities covers anything related

to coding up a new algorithm, consulting with my company, or jotting down a new research concept on a scrap of paper.

What is the most interesting project you are currently involved with?

The most interesting project that I currently lead involves an investigation in developing robot therapy interventions for young children with motor disabilities. For this project, we have developed an interactive therapy game called SuperPop VR that requires children to play within a virtual environment based on a therapist-designed protocol. A robot playmate interacts with each child during game play and provides both corrective and motivational feedback. An example of corrective feedback is when the robot physically shows the child how to interact with the game at the correct movement speed (as compared to a normative data profile). An example of motivational feedback is when the robot, through social interaction, encourages the child when they have accomplished their therapy exercise goal. We've currently deployed the system in pilot studies with children with Cerebral Palsy and have shown positive changes with respect to their kinematic outcome metrics. We're pushing the state-of-the-art in this space by incorporating additional factors for enhancing the long-term engagement through adaptation of both the therapy protocol as well as the robot behaviors.

How do you balance being involved in so many different aspects of the AI community?

In order for me to become involved in any new AI initiative and still maintain a healthy work-life balance, I ask myself: Is this initiative something that's important to me and aligned with my value system; Can I provide a unique perspective to this initiative that would help to make a difference; Is it as important or more important than other initiatives I'm involved in; and Is there a current activity that I can replace so I have time to commit to the initiative now or in the near-future. If the answer is yes to all those questions, then I'm usually able to find an optimal balance of involvement in the different AI initiatives of interest.

What is your favorite CS or AI-related movie or book and why?

My favorite AI-related movie is *The Matrix*. What fascinates me about *The Matrix* is the symbiotic relationship that exists between humans and intelligent agents (both virtual and physical). One entity can not seem to exist without the other. And operating in the physical world is much more difficult than operating in the virtual, although most agents don't realize that difference until they accept the decision to navigate in both types of worlds.

Help us determine who should be in the AI Matters spotlight!

If you have suggestions for who we should profile next, please feel free to contact us via email at aimatters@sigai.acm.org.

AI Buzzwords Explained: Distributed Constraint Optimization Problems

Ferdinando Fioretto (University of Michigan; fioretto@umich.edu)
William Yeoh (Washington University in St. Louis; wyeoh@wustl.edu)
DOI: 10.1145/3175502.3175506

The power network is the largest operating *machine* on earth, generating more than US$400bn a year[1] keeping the lights on for our homes, offices, and factories. A significant concern in power networks is for the energy providers to be able to generate enough power to supply the demands at any point in time. Short terms demand peaks are however hard to predict and, thus, in the modern *smart electricity grid*, the energy providers can exploit the demand-side flexibility of the consumers to reduce the peaks in load demand.

This control mechanism is called *Demand-side management* (DSM). DSM can be obtained by scheduling *shiftable loads* (i.e., a portion of power consumption that can be moved from a time slot to another) from peak to off-peak hours (Fioretto, Yeoh, & Pontelli, 2017; Logenthiran, Srinivasan, & Shun, 2012; Voice, Vytelingum, Ramchurn, Rogers, & Jennings, 2011). In a simplified version of this problem, the energy provider has a desired maximal amount of power that it can generate and, thus, use to serve its customers. When the predicted amount of customer loads exceed such amount, the provider has to reschedule some of these loads in different time slots to satisfy the constraint on the maximum power capacity.

Such an approach, however, requires the provider to control a portion of the consumer's electrical appliances, affecting privacy and users' autonomy. On the other hand, residential and commercial buildings are progressively being partially automated, through the introduction of smart devices (e.g., smart thermostats, circulator heating, washing machines). Household penetration is at 5.8% in 2016 and is expected to hit 18.6% in 2020.[2]

[1] U.S. Energy Information Administration

[2] https://www.statista.com/outlook/279/109/smart-home/united-states#market-driver

Device scheduling can be executed by the users, without the control of a centralized authority, and a coordinated device scheduling within a neighborhood of buildings can be used as a DSM strategy, preserving user data privacy. Figure 1 illustrates such scenario.

One possible way to solve this problem is through the use of *Distributed Constraint Optimization Problems* (DCOPs) (Fioretto, Pontelli, & Yeoh, 2016; Modi, Shen, Tambe, & Yokoo, 2005; Petcu & Faltings, 2005a). DCOP algorithms are a class of distributed cooperative multi-agent algorithms in which several autonomous agents coordinate their decisions to achieve a shared goal while accounting for personal preferences. The agents can be thought of as software programs whose execution does not depend on the execution of other agents. Their actions are expressed using the concept of *variables*, i.e., abstract entities that can take one out of several values (describing the possible set of actions for the agent). Each agent needs to decide which value to assign to its variables. The outcome of an action is expressed in terms of a *cost* (or *reward*) and typically depends on the joint action of multiple agents. The goal of a DCOP is expressed in the form of an objective function to be minimized (or maximized). To coordinate their actions, agents employ a message passing mechanism realized through a networked communication.

Mathematically, a DCOP is composed by the following entities:

- $\mathcal{A} = \{a_1, \dots, a_p\}$: The set of autonomous agents participating in the problem.
- $\mathcal{X} = \{x_1, \dots, x_n\}$: The set of variables in the problem. Each variable is controlled by exactly one agent.
- $\mathcal{D} = \{D_1, \dots, D_n\}$: The domains for the variables in $\mathcal{X}$, where D_i represents the set of possible values that the variable x_i may be assigned.

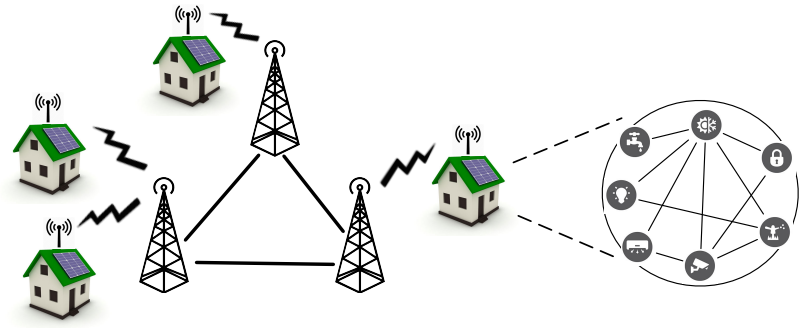


Figure 1: An illustration of a smart neighborhood with each home controlling a set of smart devices.

- $\mathcal{C} = \{c_1, \ldots, c_e\}$: The set of problem constraints. Each constraint c_i is a function that involves (multiple) variable(s) from $\mathcal{X}$ and associates a cost for each combination of their value assignments.
- $\alpha : \mathcal{X} \rightarrow \mathcal{A}$: A mapping that associates variables to agents, expressing which agent controls which variables.

The goal of the problem is to find an assignment for the agent variables that minimizes the sum of all costs over all constraints. Since the agents are physically distributed across a network, all communication take the form of messages. Thus, agents coordinate the value assignment for their variables following a given distributed protocol. In addition, agents knowledge is limited to their resources: each agent knows exclusively the outcomes of the variables it controls and the constraints it shares with some other agents. This scheme is effective to design algorithms that preserve data privacy.

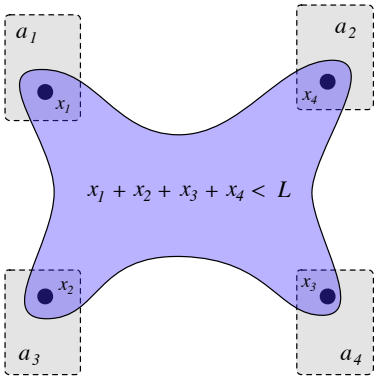


Figure 2: An illustration of a smart neighbor with each home controlling a set of smart devices.

In the DSM smart device scheduling example in Figure 2, we illustrate a smart neighborhood composed of 4 homes, each represented by one agent: $a_1, \ldots, a_4$. In this simplified scenario, each agent controls one single variable $x_1, \ldots, x_4$, which represents a simplified electrical appliance that can be switched *off* (consuming 0 kWh) or *on* (consuming 2 kWh). Assume the energy provider imposes a total consumption limit of 4 kWh. We represent the domains of each variable with the set $\{0, 2\}$, indicating the consumption associated with the device's actions. The only constraint of the problem is expressed with the formula $x_1 + x_2 + x_3 + x_4 \leq 4$, meaning that the aggregated energy consumption cannot exceed 4 kWh. A possible solution to the problem is thus having agents a_1, and a_2 switching their appliances *on*, and thus consuming a total of $x_1 = 2 + x_2 = 2 = 4$ kWh, and agents a_3 and a_4 switching their appliance *off*, thus consuming 0 kWh. A more detailed description of the DSM scheduling application and its corresponding DCOP model can be found in (Fioretto et al., 2017; Tabakhi, Le, Fioretto, & Yeoh, 2017).

The DCOP framework is general and offers a flexible tool to model a wide variety of problems. Examples of use of DCOPs to solve distributed problems include *service-oriented computing*, that relies on sharing resources over a network, focusing on maximizing the effectiveness of the shared resources which are used by multiple applications (Choudhury, Dey, Dutta, & Choudhury, 2014; Jin, Cao, & Li, 2011; Li, Wang, Ding, & Li, 2014), *sensor network problems*, which consist of coordinating a large number of inexpensive and autonomous sensor nodes, constrained by a limited communication range and battery

life (Hosseini Semnani & Basir, 2013; Ota, Matsui, & Matsuo, 2009; Stranders, Farinelli, Rogers, & Jennings, 2009; Zhang, Wang, Xing, & Wittenberg, 2005), and many others (Brys, Pham, & Taylor, 2014; Gaudreault, Frayret, & Pesant, 2009; Junges & Bazzan, 2008; Kumar, Faltings, & Petcu, 2009; Miller, Ramchurn, & Rogers, 2012; Rust, Picard, & Ramparany, 2016; Yeoh & Yokoo, 2012; Zivan, Yedidsion, Okamoto, Glinton, & Sycara, 2015). More examples can be found in a recent survey (Fioretto, Pontelli, & Yeoh, 2016).

It turns out that it is difficult (NP-hard) to optimally solve this kind of problems, and that there is a close relationship between the amount of information that needs to be encoded in a message (message size) vs. the number of messages exchanged by the agents (network load). Thus, an extensive piece of the DCOP literature focuses on the study of algorithms that trade off solution quality for faster runtime and reduced use of network resources (Farinelli, Rogers, Petcu, & Jennings, 2008; Fioretto, Yeoh, & Pontelli, 2016; Maheswaran, Pearce, & Tambe, 2004; Nguyen, Yeoh, & Lau, 2013; Ottens, Dimitrakakis, & Faltings, 2017; Pearce & Tambe, 2007; Petcu & Faltings, 2007a; Yeoh, Sun, & Koenig, 2009; Zhang et al., 2005).

As one of the motivations for the use of DCOPs is the preservation of privacy, there is also a large body of work on privacy-preserving algorithms (Grinshpoun & Tassa, 2016; Léauté & Faltings, 2011a, 2013; Tassa, Grinshpoun, & Zivan, 2017; Tassa, Zivan, & Grinshpoun, 2016).

Additionally, the DCOP model has also been extended to handle problems where agents have multiple objectives (Delle Fave, Stranders, Rogers, & Jennings, 2011; Matsui, Silaghi, Hirayama, Yokoo, & Matsuo, 2012), problems of a dynamic nature (i.e., where the problem changes over time) (Hoang et al., 2016, 2017; Nguyen, Yeoh, Lau, Zilberstein, & Zhang, 2014; Petcu & Faltings, 2005b, 2007b; Yeoh, Varakantham, Sun, & Koenig, 2015), and problems with uncertainty (i.e., where constraint costs depend from uncertain factors, such as weather) (Atlas & Decker, 2010; Le, Fioretto, Yeoh, Son, & Pontelli, 2016; Léauté & Faltings, 2011b; Nguyen, Yeoh, & Lau, 2012; Stranders, Delle Fave, Rogers, & Jennings, 2011).

A forum for discussion on DCOP algorithms and applications has been held at the *Optimization in Multi-Agent Systems* (OptMAS) workshop since 2010 at the *International Conference on Autonomous Agents and Multiagent Systems* (AAMAS). Recent dissertations within the past 5 years include (Billiau, 2015; Delle Fave, 2012; Fioretto, 2016; Grubshtein, 2012; Guttierez, 2012; Hanada, 2017; Hatano, 2013; Kim, 2015; Miller, 2014; Netzer, 2015; Okimoto, 2012; Ottens, 2012; Ueda, 2014; Yedidsion, 2015).

References

Atlas, J., & Decker, K. (2010). Coordination for uncertain outcomes using distributed neighbor exchange. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 1047–1054).

Billiau, G. (2015). *An equitable approach of solving distributed constraint optimization problems* (PhD thesis). University of Wollongong, Wollongong (Australia).

Brys, T., Pham, T. T., & Taylor, M. E. (2014). Distributed learning and multi-objectivity in traffic light control. *Connection Science*, *26*(1), 65–83.

Choudhury, B., Dey, P., Dutta, A., & Choudhury, S. (2014). A multi-agent based optimised server selection scheme for SOC in pervasive environment. In *Advances in practical applications of heterogeneous multi-agent systems* (pp. 50–61).

Delle Fave, F. M. (2012). *Theory and practice of coordination algorithms exploiting the generalised distributive law* (PhD thesis). University of Southampton, Southampton (UK).

Delle Fave, F. M., Stranders, R., Rogers, A., & Jennings, N. (2011). Bounded decentralised coordination over multiple objectives. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 371–378).

Farinelli, A., Rogers, A., Petcu, A., & Jennings, N. (2008). Decentralised coordination of low-power embedded devices using the Max-Sum algorithm. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 639–646).

Fioretto, F. (2016). *Exploiting the structure of distributed constraint optimization problems with applications in smart grids* (PhD thesis). New Mexico State University, Las Cruces (United States).

Fioretto, F., Pontelli, E., & Yeoh, W. (2016). Distributed constraint optimization problems and applications: A survey. *CoRR*, *abs/1602.06347*.

Fioretto, F., Yeoh, W., & Pontelli, E. (2016). Multi-variable agent decomposition for DCOPs. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 2480–2486).

Fioretto, F., Yeoh, W., & Pontelli, E. (2017). A multiagent system approach to scheduling devices in smart homes. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 981–989).

Gaudreault, J., Frayret, J.-M., & Pesant, G. (2009). Distributed search for supply chain coordination. *Computers in Industry*, *60*(6), 441–451.

Grinshpoun, T., & Tassa, T. (2016). P-SyncBB: A privacy preserving branch and bound DCOP algorithm. *Journal of Artificial Intelligence Research*, *57*, 621–660.

Grubshtein, A. (2012). *Distribute search by agents with personal preferences* (PhD thesis). Ben-Gurion University of the Negev, Beer-Sheva (Israel).

Guttierez, P. (2012). *Distributed constraint optimization related with soft arc consistency* (PhD thesis). Universitat Autònoma de Barcelona, Bellaterra (Spain).

Hanada, K. (2017). *Distributed lagrangian relaxation protocol and its applications* (PhD thesis). Kobe University, Kobe (Japan).

Hatano, D. (2013). *Solutions for constraint problem under the dynamic/distributed environment* (PhD thesis). Kobe University, Kobe (Japan).

Hoang, K. D., Fioretto, F., Hou, P., Yokoo, M., Yeoh, W., & Zivan, R. (2016). Proactive dynamic distributed constraint optimization. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 597–605).

Hoang, K. D., Hou, P., Fioretto, F., Yeoh, W., Zivan, R., & Yokoo, M. (2017). Infinite-horizon proactive dynamic DCOPs. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 212–220).

Hosseini Semnani, S., & Basir, O. A. (2013). Target to sensor allocation: A hierarchical dynamic distributed constraint optimization approach. *Computer Communications*, *36*(9), 1024–1038.

Jin, Z., Cao, J., & Li, M. (2011). A distributed application component placement approach for cloud computing environment. In *Proceedings of the international conference on dependable, autonomic and secure computing* (pp. 488–495).

Junges, R., & Bazzan, A. L. (2008). Evaluating the performance of DCOP algorithms in a real world, dynamic problem. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 599–606).

Kim, Y. (2015). *Application of techniques for MAP estimation to distributed constraint optimization problem* (PhD thesis). University of Massachusetts at Amherst, Amherst (United States).

Kumar, A., Faltings, B., & Petcu, A. (2009). Distributed constraint optimization with structured resource constraints. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 923–930).

Le, T., Fioretto, F., Yeoh, W., Son, T. C., & Pontelli, E. (2016). ER-DCOPs: A framework for distributed constraint optimization with uncertainty in constraint utilities. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 606–614).

Léauté, T., & Faltings, B. (2011a). Coordinating logistics operations with privacy guarantees. In *Proceedings of the international joint conference on artificial intelligence (ijcai)* (pp. 2482–2487).

Léauté, T., & Faltings, B. (2011b). Distributed constraint optimization under stochastic uncertainty. In *Proceedings of the aaai conference on artificial intelligence (aaai)* (pp. 68–73).

Léauté, T., & Faltings, B. (2013). Protecting privacy through distributed computation in multi-agent decision making. *Journal of Artificial Intelligence Research*, *47*, 649–695.

Li, X., Wang, H., Ding, B., & Li, X. (2014). MABP: an optimal resource allocation approach in data center networks. *Science China Information Sciences*, *57*(10), 1–16.

Logenthiran, T., Srinivasan, D., & Shun, T. (2012). Demand side management in smart grid using heuristic optimization. *IEEE Transactions on Smart Grid*, *3*(3), 1244–1252.

Maheswaran, R., Pearce, J., & Tambe, M. (2004). Distributed algorithms for DCOP: A graphical game-based approach. In *Proceedings of the international conference on parallel and distributed computing systems* (pp. 432–439).

Matsui, T., Silaghi, M., Hirayama, K., Yokoo, M., & Matsuo, H. (2012). Distributed search method with bounded cost vectors on multiple objective DCOPs. In *Proceedings of the principles and practice of multi-agent systems* (pp. 137–152).

Miller, S. (2014). *Decentralised coordination of smart distribution networks using message passing* (PhD thesis). University of Southampton, Southampton (UK).

Miller, S., Ramchurn, S. D., & Rogers, A. (2012). Optimal Decentralised Dispatch of Embedded Generation in the Smart Grid. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 281–288).

Modi, P., Shen, W.-M., Tambe, M., & Yokoo, M. (2005). ADOPT: Asynchronous distributed constraint optimization with quality guarantees. *Artificial Intelligence*, *161*(1–2), 149–180.

Netzer, A. (2015). *Distributed constraint optimization, from utilitarianism to fairness* (PhD thesis). Ben-Gurion University of the Negev, Beer-Sheva (Israel).

Nguyen, D. T., Yeoh, W., & Lau, H. C. (2012). Stochastic Dominance in Stochastic DCOPs for Risk-sensitive Applications. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 257–264).

Nguyen, D. T., Yeoh, W., & Lau, H. C. (2013). Distributed Gibbs: A memory-bounded sampling-based DCOP algorithm. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 167–174).

Nguyen, D. T., Yeoh, W., Lau, H. C., Zilberstein, S., & Zhang, C. (2014). Decentralized multi-agent reinforcement learning in average-reward dynamic DCOPs. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 1447–1455).

Okimoto, T. (2012). *Graph-based algorithms for distributed constraint satisfaction/optimization problems* (PhD thesis). Kyushu University, Fukuoka (Japan).

Ota, K., Matsui, T., & Matsuo, H. (2009). Layered distributed constraint optimization problem for resource allocation problem in distributed sensor networks. In *Proceedings of the principles and practice of multi-agent systems* (pp. 245–260).

Ottens, B. (2012). *Coordination and sampling in distributed constraint optimization* (PhD thesis). Ecole Polytechnique Fédérale de Lausanne, Lausanne (Switzerland).

Ottens, B., Dimitrakakis, C., & Faltings, B. (2017). DUCT: An upper confidence bound approach to distributed constraint optimization problems. *ACM Transactions on Intelligent Systems and Technology*, *8*(5), 69:1–69:27.

Pearce, J., & Tambe, M. (2007). Quality guarantees on k-optimal solutions for distributed constraint optimization problems. In *Proceedings of the international joint conference on artificial intelligence* (pp. 1446–1451).

Petcu, A., & Faltings, B. (2005a). A scalable method for multiagent constraint optimization. In *Proceedings of the international joint conference on artificial intelligence* (pp. 1413–1420).

Petcu, A., & Faltings, B. (2005b). Superstabilizing, Fault-Containing Distributed Combinatorial Optimization. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 449–454).

Petcu, A., & Faltings, B. (2007a). MB-DPOP: A new memory-bounded algorithm for distributed optimization. In *Proceedings of the international joint conference on artificial intelligence* (pp. 1452–1457).

Petcu, A., & Faltings, B. (2007b). Optimal solution stability in dynamic, distributed constraint optimization. In *Proceedings of the international conference on intelligent agent technology* (pp. 321–327).

Rust, P., Picard, G., & Ramparany, F. (2016). Using message-passing DCOP algo-

rithms to solve energy-efficient smart environment configuration problems. In *Proceedings of the international joint conference on artificial intelligence* (pp. 468–474).

Stranders, R., Delle Fave, F. M., Rogers, A., & Jennings, N. (2011). *U-GDL: A decentralised algorithm for dcops with uncertainty* (Tech. Rep.). Department of Electronics and Computer Science: University of Southampton.

Stranders, R., Farinelli, A., Rogers, A., & Jennings, N. R. (2009). Decentralised coordination of continuously valued control parameters using the max-sum algorithm. In *Proceedings of the international conference on autonomous agents and multiagent systems* (pp. 601–608).

Tabakhi, A. M., Le, T., Fioretto, F., & Yeoh, W. (2017). Preference elicitation for DCOPs. In *Proceedings of the international conference on principles and practice of constraint programming* (pp. 278–296).

Tassa, T., Grinshpoun, T., & Zivan, R. (2017). Privacy preserving implementation of the Max-Sum algorithm and its variants. *Journal of Artificial Intelligence Research*, *59*, 311–349.

Tassa, T., Zivan, R., & Grinshpoun, T. (2016). Preserving privacy in region optimal DCOP algorithms. In *Proceedings of the international joint conference on artificial intelligence* (pp. 496–502).

Ueda, S. (2014). *Computational coalition formation: Compact representation and constrained matching* (PhD thesis). Kyushu University, Fukuoka (Japan).

Voice, T., Vytelingum, P., Ramchurn, S., Rogers, A., & Jennings, N. (2011). Decentralised control of micro-storage in the smart grid. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 1421–1427).

Yedidsion, H. (2015). *Distributed constraint optimization for teams of mobile agents* (PhD thesis). Ben-Gurion University of the Negev, Beer-Sheva (Israel).

Yeoh, W., Sun, X., & Koenig, S. (2009). Trading off solution quality for faster computation in DCOP search algorithms. In *Proceedings of the international joint conference on artificial intelligence* (pp. 354–360).

Yeoh, W., Varakantham, P., Sun, X., & Koenig, S. (2015). Incremental DCOP search algorithms for solving dynamic DCOP problems. In *Proceedings of the international conference on intelligent agent technology* (pp. 257–263).

Yeoh, W., & Yokoo, M. (2012). Distributed problem solving. *AI Magazine*, *33*(3), 53–65.

Zhang, W., Wang, G., Xing, Z., & Wittenberg, L. (2005). Distributed stochastic search and distributed breakout: Properties, comparison and applications to constraint optimization problems in sensor networks. *Artificial Intelligence*, *161*(1–2), 55–87.

Zivan, R., Yedidsion, H., Okamoto, S., Glinton, R., & Sycara, K. (2015). Distributed constraint optimization for teams of mobile sensing agents. *Journal of Autonomous Agents and Multi-Agent Systems*, *29*(3), 495–536.

Ferdinando Fioretto is a postdoctoral researcher at the Industrial and Operations Engineering Department of the University of Michigan. His research focuses on multi-agent systems, data privacy, and discrete optimization. His dissertation was awarded the best dissertation in Artificial Intelligence from the Italian Association of Artificial Intelligence, in 2017. Additional information can be found at: http://www-personal.umich.edu/~fioretto/.

William Yeoh is an assistant professor in the Computer Science and Engineering Department at Washington University in St. Louis. His research interests include multi-agent systems, distributed constraint reasoning, heuristic search, and planning with uncertainty. He is an NSF CAREER awardee and was named in IEEE's AI's 10-to-Watch list in 2015. Additional information can be found at: https://sites.wustl.edu/wyeoh/.

Intelligent Workflows for Visual Stylometry

Catherine A. Buell (Fitchburg State University; cbuell1@fitchburgstate.edu)
Yolanda Gil (USC Information Sciences Institute; gil@isi.edu)
William P. Seeley (Boston College; seelyw@bc.edu)
Ricky J. Sethi (Fitchburg State University; rickys@sethi.org)
DOI: 10.1145/3175502.3175507

Using Intelligent Workflows to Analyze Artistic Style

USC Information Sciences Institute (ISI) alumnus Ricky J. Sethi and his colleagues at Fitchburg State University in Massachusetts are using ISI's WINGS workflow system for art history in the WAIVS (Workflows for Analysis of Images and Visual Stylometry) project. WAIVS workflows were demonstrated at a workshop held at the Fitchburg Art Museum (FAM) in Spring, 2017.

The WAIVS project is funded by a grant from the National Endowment for the Humanities (NEH). The principal investigators are Sethi, assistant professor in the Computer Science Department at Fitchburg State, and colleagues Catherine A. Buell, assistant professor in the Mathematics Department, and William P. Seeley, a visiting scholar in the Department of Psychology at Boston College. Other project members include RaghuRam Rangaraju and Jake Lee, both computer science students at Fitchburg State. The project is in collaboration with Dr. Mary M. Tinti of the Fitchburg Art Museum, Dr. Yolanda Gil of the USC Information Sciences Institute, and Dr. Charlene Villaseñor Black of the department of art history at UCLA.

The focus of the WAIVS project is in visual stylometry, an emerging field that applies image analysis and machine learning tools to digital artwork for art analysis and investigation. Visual Stylometry combines research and methods from art history, computer science, and cognitive science to help quantify the style of an artist. It can be used to provide clues to the visual elements of a painting that enable viewers to categorize works as belonging to different artistic styles and can contribute to an analysis of the qualities of an artwork that affect how we experience it.

Figure 1: The WAIVS Group: WAIVS project members, from left to right: RaghuRam Rangaraju a graduate student at Fitchburg State University; Ricky Sethi, assistant professor of computer science at Fitchburg State University; Catherine Buell, assistant professor of mathematics at Fitchburg State University; William Seeley, visiting scholar at Boston College; and Jake Lee, undergraduate student at Fitchburg State University. On the far right is WAIVS collaborator Charlene Villaseñor-Black, a professor of art history at UCLA.

Instead of relying only on what our senses perceive, we can come up with artistically relevant computational features and techniques to quantify and compare aspects of artistic style over the course of the career of an individual artist, among artists who share in a common artistic style, and across different schools of art. Although there have been tremendous advances in the field of image processing that are relevant to visual stylometry, they are not very accessible to art historians. They have yet to be translated into a medium that is accessible to researchers in arts related fields from psychology of art to art history.

To address this, WAIVS is using workflows to provide an accessible visual programming interface that simply shows how the data is generated and used by different computational steps. Workflows effectively capture complex

multi-step data analysis methods in a simple dataflow graph. WAIVS builds upon the WINGS workflow system, developed by Gil's group at ISI, because it adds intelligent reasoning to workflows.

It uses a unique workflow system that uses artificial intelligence planning techniques and semantic web languages to capture expert knowledge about setting up the parameters that control the image analysis algorithms, so that users can get recommendations of parameter settings to create valid workflows that work best with their data.

Sethi, who is an expert in video processing, developed workflows that include state-of-the-art methods such as deep learning and convolutional neural networks to analyze images. Sethi's postdoctoral research at ISI was under a prestigious NSF Computing Innovation Fellows (CIFellows) award. During that time, he collaborated with Gil on combining text and image analysis workflows to detect human trafficking by analyzing personal ads in Web sites. They recently published a paper about the use of deep learning techniques in workflows to capture artistic style, which will appear in the Future Generation Computer Systems journal.

Using WINGS, WAIVS image processing experts create workflows that capture state-of-the-art image processing techniques. Current workflows created by the WAIVS group include entropy calculation, discrete tonal analysis, and convolutional neural networks. Art historians learned to use these workflows during the workshop.

The 2017 WAIVS workshop was attended by more than twenty art historians, mostly in the New England area, and was supported by the American Society for Aesthetics and the New England Museum Association. The workshop was held in the room that hosts the exhibit of Lionel Reinford, a well-known local painter. The discussions centered on possible approaches to quantifying artistic style. As Sethi, Buell, and other WAIVS project members demonstrated workflows to compute the entropy of a painting and other quantitative ways to represent a painting, art historians discussed the possibilities of using such measures to design more formal descriptions of artistic style.

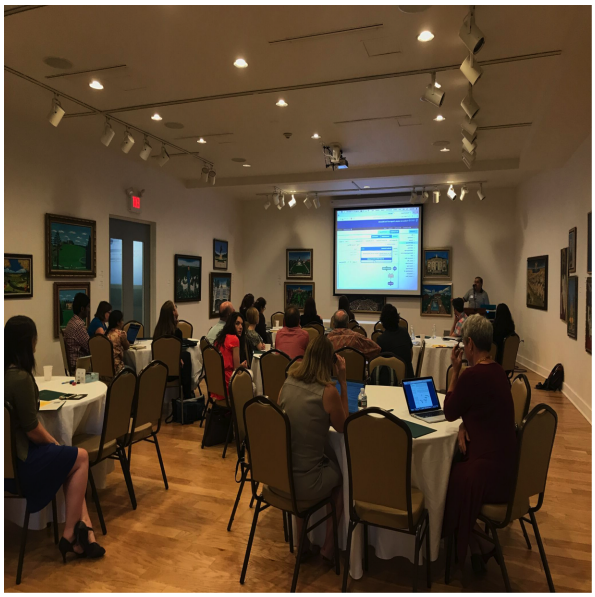


Figure 2: Presentation by Sethi on overview of WINGS: Sethi gave a presentation of WINGS workflows to analyze artistic style. Sethi, who was a postdoctoral researcher at ISI under a prestigious CRA CIFellows scholarship, is an expert in video and image processing. The workshop was held with a backdrop of the exhibit of Lionel Reinford, a well-known local painter.

The first talk was by Daniel Graham of Hobart and William Smith College, who discussed the neurobiological aspects of artistic style. Next, Gil gave an introduction to workflows and to the WINGS intelligent workflow system.

Seeley discussed the origins of the WAIVS project as a collaborative teaching exercise with Buell at Bates College. The goal of the initial project was to foster interdisciplinary collaboration among undergraduates in humanities, mathematics, and computer science. Seeley mentioned that the initial choice of focus on Hudson River School and Impressionist landscape paintings was strategic.

The particular Hudson River school landscape images in the set were chosen because they share a similar general composition and palette that can be traced to earlier seventeenth, eighteenth, and nineteenth century Dutch and English landscapes. This ties the work to E. H. Gombrich's research on the development of artistic style. Further, all of the works chosen are in the public domain and available via online archives like WikiArt. These works represent styles that are familiar

Figure 3: Presentation by Buell on WINGS experiments: Catherine Buell of the Mathematics Department at Fitchburg State University shows workshop participants how a convolutional neural network learns from examples of Van Gogh's artwork about his artistic style (left of her slide), and can then render any image (middle of her slide) using the distinctive strokes and colors of the Dutch painter (right of her slide).

and well represented in art museums.

This makes WAIVS accessible as a teaching exercise for students, researchers and the broader public. Finally, the choice of paintings with similar palettes and composition, as well as the choice to contrast Hudson River School and Impressionist paintings, was designed to test an initial hypothesis that texture information, which is indicative of differences in brushstroke styles, would be sufficient to classify artworks by school and individual artists.

Sethi, Buell, and their students gave a demonstration of the WAIVS system, and guided participants through several practical exercises to use WAIVS workflows to analyze a variety of paintings, some of them from a current exhibit at the host museum.

Some WAIVS workflows capture interesting quantitative measures of an image's characteristics. For example, one of the workflows generates an entropy value and an entropy image, allowing art historians to compare different paintings in terms of their entropy levels.

Another workflow uses a convolutional neural network, and is trained with examples of a painter's artwork (the style image) to then render any image (the content image) using the

Figure 4: Workflow for calculating entropy: A workflow that generates an entropy value and an entropy image Church's The Heart of the Andes (1859)

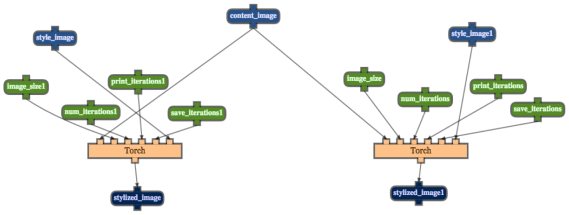


Figure 5: Workflow for Generating Stylized Images: the workflow GenerateStylizedImages uses convolutional neural networks to process two separate paintings (the style images) and renders an image (the content image) in the style of those paintings. Comparing the two resulting synthetic stylized images helps art historians contrast the styles of the paintings

distinctive strokes and colors of that painter. This is based on a technique developed by Leon Gatys, Alexander Ecker, and Matthias Bethge from Tübingen in Germany in 2015. The WINGS workflow was implemented using the Torch open source software for deep learning. The components of these workflows can be linked together to create different analyses.

Workshop participants worked with images by contemporary painter Shelley Reed, the subject of a current exhibit at the museum. Reed appropriates imagery from seventeenth, eighteenth, and nineteenth century Northern European painters in her works. Workshop participants learned how to use the WAIVS software to evaluate differences in artistic style between Reed's paintings and the earlier paintings.

Participants used the GenerateStylizedImages workflow with the Cropped grayscale versions of A) Edwin Landseer's Portrait of Mr. Van Amburgh, as He Appeared with His Animals at the London Theatres (1847) and B) Shelley Reed's Tiger (after Landseer and

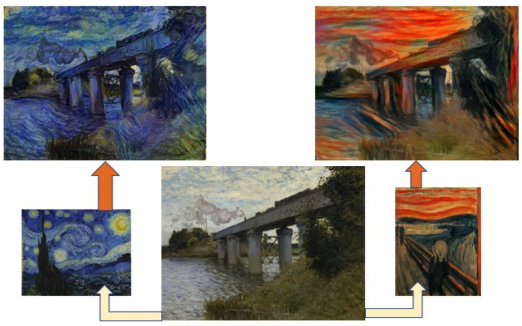


Figure 6: Example of Stylized Image Formation. An illustration of how convolutional neural networks in the GenerateStylizedImages workflow are used to transform Monet's The Railroad bridge in Argenteuil (1873) (bottom middle) into the painterly styles of Munch's The Scream (right) and Van Gogh's The Starry Night (left). The resulting synthetic stylized images are shown at the top.

Figure 7: More Generalized Stylized Images: Workshop participants used the GenerateStylizedImages workflow to compare a painting from Shelley Reed (A) with a painting by Edwin Landseer (B) that she appropriated in her painting, both used as input style images. The content image from Frederick Church (bottom) was used to generate two synthetic stylized images (top), which expose stylistic differences in the strokes of the tiger's stripes were painted in the Reed and the original Landseer paintings. The starker tonal contrasts of the Reed painting are also evident in the way the waterfall and the sky have been depicted in the two stylized images.

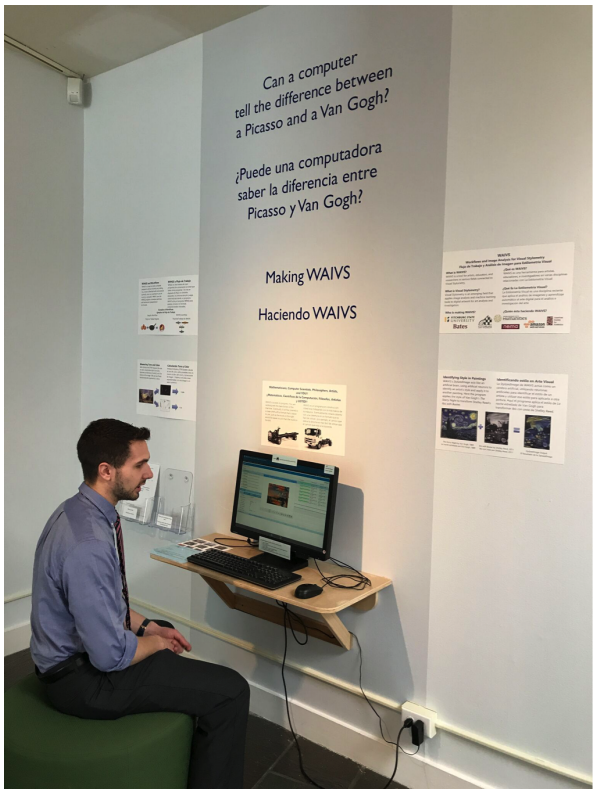


Figure 8: Student demonstrating WAIVS: Fitchburg State University student Jake Lee shows the "Making WAIVS exhibit" at the Fitchburg Art Museum in Fitchburg, Massachusetts. Museum goers interact with workflows to analyze paintings.

Thiele) (2007) as the style images. Frederick Church's Heart of the Andes (1859) was used as the content image. The resulting synthetic images, A) Stylized Landseer and B) Stylized Reed, reflect several interesting stylistic differences between the original Landseer and Reed paintings.

The most striking can be seen in the grove of trees in the foreground right of the paintings. The trees are rendered in more tightly packed and sharply articulated stripes in the Stylized Reed than the Stylized Landseer. This difference recapitulates differences in the way that the tiger's stripes were painted in the Reed and the original Landseer paintings. The starker tonal contrasts of the Reed painting are also evident in the way the waterfall and the sky have been depicted in the two stylized images.

Workshop attendees also had the opportunity to examine Reed's artistic style in the exhibit

Figure 9: Participants at WAIVS: Workshop participant John Kulvicki, from the Department of Philosophy at Dartmouth, analyzed paintings with WINGS workflows on a mobile phone. He is interested in understanding the subjectivity of artistic style.

Curious Nature, running at the Fitchburg Art Museum (February 12 - June 4). A demonstration version of WAIVS is currently available for use by the general public in association with the Reed exhibition. The exhibit materials are also offered in Spanish to appeal to the local Latino population.

Workshop participant John Garton of Clark University proposed using workflows to understand the 3D effect on color when paintings have texture that changes how the color is reflected on the 3D structure. He explained how El Greco used lapis in the mixes he did for blues, giving his paintings unique color effects. Workshop participants Valerie Kinkade of the Museum and Collector Resource and Amy Schlegel discussed how art historians collect mass spectrometry to understand the chemical composition of the pigments, as well as stratigraphy data about the paint thickness and its 3D structure. This kind of data opens the door to new research to analyze those kinds of artistic elements and the 3D effects on the perception of color in paintings. Kinkade also saw applications in legal aspects of copyright infringement of paintings.

Charlene Villaseñor-Black, a professor in the department of Art History at the University of California Los Angeles (UCLA), discussed early uses of technology as a tool by painters, exemplified by the use of camera obscura by

Figure 10: Student helping participants at WAIVS: Workshop participants John Garton of Clark University, Amy Schlegel, and Valerie Kinkade of Museum and Collector Resource, run workflows to analyze the artistic style of Albert Bierstadt's 1895 painting "The Morteratsch Glacier, Upper Engadine Valley, Pontresina," as WAIVS student Jake Lee looks on.

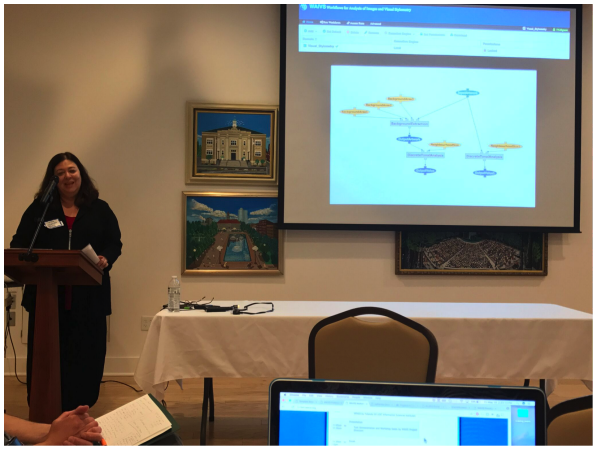


Figure 11: Villaseñor presenting at WAIVS: Charlene Villaseñor-Black, a professor of Art History at UCLA, presented examples of early uses of technology as a tool by painters, and discussed the potential of workflows and computer vision tools to help art historians think differently about style, and to open the doors for students to learn about visual style in a more analytical way.

Vermeer and Caravaggio, and the different levels of detail designed to reflect the eye's perception in the forefront figures of Las Meninas from Velazquez. She discussed the potential of workflows and computer vision tools to help art historians think differently about style, and to open the doors for students to learn about visual style in a more analytical way.

Sethi was particularly proud to see this workshop come together. "My wife is a historian, and I see first hand how challenging it is for people in the humanities to access the powerful technologies for data science that are available today. Ever since I started to use WINGS at ISI, I could see that workflows can be a game changer for historians. For art historians in particular, workflows can bring very sophisticated tools from image processing into their hands, and allow them to experiment with different mathematical measures of the properties of an image that they could then ascribe to artistic style."

Gil, who uses WINGS workflows to teach data science to non-computer science students at USC, was not surprised that the art historians were able to run sophisticated quantitative analyses on paintings. "What is unique about the WAIVS project is the use of methods from computer vision in order to give quantitative definitions of technical terms in art history," she said. "This project is visionary in bringing recent revolutionary deep learning AI techniques to quantify the study of art, and putting them squarely in the hands of humanities researchers."

"I am impressed by WAIVS and its potential to revolutionize the way we look, the way we think, the way we see images" Villaseñor-Black underscored. "The WAIVS tool is able to do things with images that art historians cannot do, such as measure entropy or remove the chromatic value from the foreground, or transfer what it calls 'style' from one image to another. These are not skills that art historians are trained in, or things we can currently do, and they have the potential to radically change how we look at and think about style."

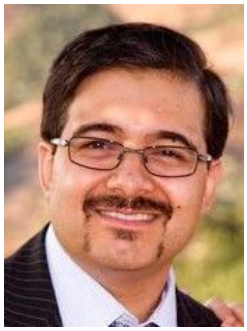
Acknowledgments

This research was supported in part by the US National Science Foundation (NSF) under grant #1019343 to the Computing Research Association for the CIFellows Project, the National Endowment for the Humanities (NEH) Grant under Award HD-248360-16, the Amazon AWS Research Grant program (AMZN), the American Society for Aesthetics (ASA), the New England Museum Association (NEMA), and the Fitchburg Art Museum (FAM).

Catherine A. Buell is an Assistant Professor in Mathematics at Fitchburg State University. She has presented her research, in algebraic groups and generalized symmetric spaces, at many national and international conferences. In addition to her theoretical work, she publishes in mathematics education and has worked since 2007 with local elementary and middle school teachers. For the past two summer, she has created educational materials for CCICADA (The Command, Control and Interoperability Center for Advanced Data Analysis) which is a US Department of Homeland Security University Center of Excellence. Catherine received her Ph.D. and M.S. in Mathematics from North Carolina State University and her B.S. in Mathematics with concentrations in dance and computer programming from Springfield College. Previously, she taught at North Carolina State University and Bates College, where besides teaching the full range of undergraduate courses, she also created the courses Applications of Abstract Algebra and Applied Linear Algebra where mathematics students utilized computation tools and delve into real-life applications of advanced theoretical mathematics and statistics. Catherine is active in the Association for Women in Mathematics (AWM) and the Mathematics Association of America (MAA) where she has hosted numerous sessions and conferences, mentored, and worked on committees.

Yolanda Gil is Associate Division Director at the Information Sciences Institute of the University of Southern California, and Research Professor in the Computer Science Department. She received her M.S. and Ph. D. degrees in Computer Science from Carnegie Mellon University. Dr. Gil leads a group that conducts research on various aspects of Interactive Knowledge Capture. Her research interests include intelligent user interfaces, knowledge-rich problem solving, scientific and grid computing, and the semantic web. An area of recent interest is large-scale distributed data analysis through semantic workflows. Dr. Gil was elected to the Council of the American Association of Artificial Intelligence (AAAI), and was program co-chair of the AAAI conference in 2006. She served in the Advisory Committee of the Computer Science and Engineering Directorate of the National Science Foundation. She currently chairs the W3C Provenance Group, an effort to chart the state-of-the-art and possible standardization efforts in this area.

William P. Seeley is a Visiting Scholar in the Department of Psychology at Boston College. He holds a Ph.D. in philosophy from CUNY-The Graduate Center, an M.F.A. in sculpture from Columbia University, and a B.A. in philosophy from Columbia University. His research interests lie at the intersection of philosophy of art, cognitive science, and embodied cognition. His research in cognitive science and aesthetics has been published in the British Journal of Aesthetics, Journal of Aesthetics and Art Criticism, Journal of Vision, Psychology of Aesthetics, Creativity and the Arts, Philosophical Psychology, and Review of Philosophy and Psychology, as well as a number of edited volumes. His welded steel constructions have been exhibited in New York City and at a number of colleges and university galleries, including a solo exhibition of outdoor works in Ezra Stiles College at Yale University. He is currently working on a book, Attentional Engines: A Perceptual Theory of the Arts.



Ricky J. Sethi is an Assistant Professor in Computer Science at Fitchburg State University and is also the Director of Research for the Madsci Network. Prior to FSU, he was a Research Scientist at UMass Amherst/UMass Medical School and at UCLA/USC Information Sciences Institute, where he was chosen as an NSF Computing Innovation Fellow (CIFellow) by the CCC and the CRA. Before that, he was a Postdoctoral Scholar at UCR, where he was the Lead Integration Scientist for the WASA project and participated in ONR's Empire Challenge 10. Ricky has authored or co-authored over 30 peer-reviewed papers, book chapters, and reports and made numerous presentations on his research in machine learning, computer vision, social computing, and data science. He has taught various courses in Computer Science, Physics, and General Science. Ricky has also supervised/mentored undergraduate students, graduate students, and postdoctoral students at UCLA, USC, and UMass. Ricky has served as a Panelist for the NSF Cyberlearning program, as an Editorial Board Member for the International Journal of Computer Vision & Signal Processing, and a Program Committee member for various conferences. In addition, he is a member of the YSP/Madsci Financial Board, a member of the American Institute of Physics, and a member of IEEE.

AI Education: Adaptive Planning

Joshua Eckroth (Stetson University; jeckroth@stetson.edu)
DOI: 10.1145/3175502.3175508

Introduction

In this column, we focus on designing assignments and projects that make use of planning engines. Planning has been one of the pillars of artificial intelligence since the origin of the field, and the research community remains active, as evidenced by competitions such as the IPC and conferences such as ICAPS.

Planning's practical applications and the increasing sophistication of planning engine design warrant its place in the classroom. However, planning's historical roots bring some negative side effects: planning is not "hot" like deep learning, thus reducing student interest; and many examples, assignments, and projects related to planning may appear stale and uninteresting since they may well be decades old.

An Adaptive Approach

In my experience, students are more engaged with an assignment or project if it seems "fresh," or stated less colloquially, adaptive to the student, local community, or zeitgeist. The material does not need to be entirely novel or make use of the most recent advances in the field. But to fully engage a student, we must find ways to position a project as tailored to the individual or class, or addressing a societal need, so that they feel there is some wider significance in completing it.

An Example: Git Planner

I have found success in engaging students by asking them to build planners for commonplace but distinctive skills and systems. For example, I designed a project called *Git Planner* (Neller et al., 2017), available online, that modeled the logic of the Git version control system and commands such as *git add*, *git revert*, etc. The planner would find a sequence of Git commands to take a repository's initial state and transform it into a user-provided goal state, e.g., that all files are committed or a certain file was reverted to a former state. My goal was to pique interest by asking students to model and automate the use of a famously difficult tool. The project was fairly narrow in its ambitions, modeling only a few common Git commands, but the students knew they were embarking on a task that had little precedent. After the course, a senior student recently completed an expanded version of the project for his year-long capstone.

A Modern Curriculum

Planning's long history has yielded a broad and mature set of techniques. A deep investigation would require at least a full course. Many schools do not have the opportunity to introduce a course focused on planning. Instead, I recommend covering a few significant topics to introduce students to planning and its applications. The idea of *planning as search* may be covered early in an AI course. Pathfinding for characters in a game is a classic example of simple planning, and can be applied to modern games to ensure student interest. Planning a sequence of actions, such as assembling a desktop computer from scratch, may be accomplished with a partial order planner. This example exercise requires that students investigate exactly how a computer can be assembled, which actions depend on which other actions, and which actions can be done in any order.

Students with a background in probability may be ready to work with Markov decision processes and partially-observable Markov decision processes (Kaelbling, Littman, & Cassandra, 1998), abbreviated POMDP, in which the result of an action is nondeterministic and the agent might not even be completely sure of its present state. Here, actions that allow the agent to learn more about the environment may be equally useful as actions that help achieve long-term goals. Relevant applications are numerous (Cassandra, 1998). An adaptive application of probabilistic plan-

ning could address the contemporary focus on cybersecurity to help design a secure network, including where to introduce a firewall, which services are only accessible by passwords, password complexity requirements to reduce the chance of brute force hacking, and the likelihood of exploiting out-of-date vulnerable software. A hacker's behavior may be simulated with a POMDP planner and the network design and policies may be improved in order to limit the hacker's probability of success.

Resources

Adaptive planning assignments and projects need not be developed from scratch. The Model AI Assignments repository, from years 2010 to 2017, includes seven projects that make use of planning. Planning competitions release challenge datasets and clearly describe the intended outcomes. Competitions are often highly motivating to students and may be considered adaptive, as described above, since they are relevant for a short period of time and students participate alongside other teams across the globe – in other words, it matters that the students are working on this project at this time, and the project requires the use of state-of-the-art techniques. However, planning competitions may be too challenging for introductory projects. Some planning engines such as Fast Downward include example problems in their documentation, which can serve as a starting point for new ideas. Finally, adaptive assignments may be found by considering an apparently disconnected field of study, such as cybersecurity, marketing, or medicine just to name a few, and attempting to solve a simplified form of their everyday concerns with planning technology.

Conclusion

Adaptive planning projects are more likely to keep students engaged because adaptive projects take advantage of a unique time, place, or other context that gives the project of sense of relevance and urgency that is typically lacking from traditional textbook curricula. Planning, like other traditional subfields of artificial intelligence, particularly need adaptive materials to compete with the ever-changing trends in the popular understanding of artificial intelligence and machine learning.

References

Cassandra, A. R. (1998). A survey of POMDP applications. In *Working notes of AAAI 1998 fall symposium on planning with partially observable markov decision processes* (Vol. 1724).

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial intelligence*, *101*(1), 99–134.

Neller, T. W., Eckroth, J., Reddy, S., Ziegler, J., Bindewald, J., Peterson, G., ... Karaman, S. (2017). Model AI assignments 2017. In *Proceedings of the seventh symposium on educational advances in artificial intelligence* (pp. 4822–4824).

Joshua Eckroth is an assistant professor at Stetson University and chief architect of i2k Connect. His research interests lie in abductive reasoning and belief revision as well as computer science pedagogy. Eckroth holds a Ph.D. in computer science from the Ohio State University. He is coeditor of AITopics.

Blue Sky Ideas in Artificial Intelligence Education from the EAAI 2017 New and Future AI Educator Program

Eric Eaton (University of Pennsylvania; eeaton@cis.upenn.edu)
Sven Koenig (University of Southern California; skoenig@usc.edu)
Claudia Schulz (TU Darmstadt; schulz@ukp.informatik.tu-darmstadt.de)
Francesco Maurelli (Jacobs University Bremen; f.maurelli@ieee.org)
John Lee (Antioch University; John@AssistiveIntelligence.com)
Joshua Eckroth (Stetson University; jeckroth@stetson.edu)
Mark Crowley (University of Waterloo; mcrowley@uwaterloo.ca)
Richard G. Freedman (University of Massachusetts Amherst; freedman@cs.umass.edu)
Rogelio E. Cardona-Rivera (University of Utah; rogelio@cs.utah.edu)
Tiago Machado (New York University; tiago.machado@nyu.edu)
Tom Williams (Tufts University; williams@cs.tufts.edu)
DOI: 10.1145/3175502.3175509

Abstract

The 7th Symposium on Educational Advances in Artificial Intelligence (EAAI'17, co-chaired by Sven Koenig and Eric Eaton) launched the EAAI New and Future AI Educator Program to support the training of early-career university faculty, secondary school faculty, and future educators (PhD candidates or postdocs who intend a career in academia). As part of the program, awardees were asked to address one of the following "blue sky" questions:

1. How could/should Artificial Intelligence (AI) courses incorporate ethics into the curriculum?
2. How could we teach AI topics at an early undergraduate or a secondary school level?
3. AI has the potential for broad impact to numerous disciplines. How could we make AI education more interdisciplinary, specifically to benefit non-engineering fields?

This paper is a collection of their responses, intended to help motivate discussion around these issues in AI education.

Bridging Across Disciplines

Claudia Schulz (TU Darmstadt)

The application of AI methods to problems such as legal decision making, language translation, or gene analysis often requires the cooperation of AI experts and subject specialists, e.g., lawyers, translators, or biologists. Their ability to communicate on a common ground is a crucial factor determining the success of the project. It is thus beneficial if both parties have a basic understanding of the subject as well as of AI methods, even before the start of a project.

Universities provide a unique opportunity to both teach students becoming AI experts some subject knowledge (e.g., biology or law) and ensure that students in non-computing subjects have a basic understanding of AI techniques. A naïve approach for achieving such interdisciplinary learning is that AI students take some first-year subject courses, and subject students some introductory AI courses. Even though this approach is easy to implement, it may not achieve the intended interdisciplinary learning benefits since the courses are not tailored towards students of a different discipline (even first-year courses often provide a detailed introduction to a specific topic instead of surveying a whole field).

We here discuss two approaches based on *peer-learning*, which provide a more beneficial interdisciplinary learning environment. They share the idea that AI and subject students learn together by teaching each other.

In the *seminar-style approach*, AI students give seminars to subject students (and vice versa). These seminars may, for example, provide an overview of AI techniques or review applications of AI methods in subject areas. This approach does not only benefit the at-

tending subject students, who acquire knowledge tailored particularly to them, but also provides valuable experience to the AI student giving the seminar in explaining AI topics to the lay audience. There is clearly a lot of variability concerning the exact setup of these seminars: they can be given by a single PhD student or by a group of undergraduates, and the attendees' background can be a mixture of subjects or a single subject (in which case the seminar will cover topics and examples related to this particular subject).

In contrast to the seminar-style approach, where the speaker teaches the audience, the *project-based approach* promises mutual teaching and learning, both in terms of knowledge and skills. In this setting, an AI student and a subject student work together on a project trying to solve a problem in the subject student's area by applying AI techniques. At the start, the subject student explains subject-specific background to the AI student, whereas the AI student teaches the subject student about possible AI techniques to be used, thus creating a mutual teaching and learning environment. During the project, students will also acquire the invaluable skills of working in an interdisciplinary team. Again, there are different setups for such projects: The problem(s) to be solved can be given by faculty or be the students' own ideas, and the project can be part of a course or an extracurricular "ideas/start-up lab".

Student-Centric Discovery

Francesco Maurelli (Jacobs Univ. Bremen)

Most approaches in university teaching are based on frontal lectures, sometimes with specific lab activities and specific homeworks. The course is divided in specific modules which are explained sequentially.

I would be interested in analysing the feasibility (and try that with a real course) of a more student-centric approach, inspired by the pedagogical Montessori method (Montessori and George 1964). Although the main focus of the method has always been on children, some of those elements have been incorporated with success in secondary-school and early-undergraduate levels.

Working with an equipped lab is fundamental for this approach. Then I would imagine that each student (or maybe each group of students) could freely decide the direction of the course, based on discovery and on what they are interested in. I have recently started a cooperation with Prof. Federico Gobbo at the University of Amsterdam, to analyse the portability of some key elements of the Montessori method into AI education of young adults (so, the target group of this call).

From one side, I am interested to see how the Montessori method applied at a later age group than usual could help the students in their personal and professional development. I strongly believe that independence and the ability of thinking, reasoning, and making informed choices are key elements of the lives of active and engaged human beings, part of the society. A teaching approach which values independent thinking seems therefore a very interesting and potentially fruitful approach, albeit maybe difficult at times.

From the other side, looking for new engaging methods of teaching AI and robotics might result in students approaching the subject with curiosity and willingness, not just because it is in the study plan. This in turn might result in more people engaged in AI and Robotics, and in more passion towards the subject. It might be perceived not just as one of many lectures, but a feel of "ownership" might push for a deeper understanding of specific subjects rather than usual frontal lectures.

Challenges in this approach would be ensuring that each student (or each group of students) progress and explore the subject within some boundaries. Also, evaluation is a very delicate subject. In the original Montessori approach there is no grading for children, but it is something usually necessary in undergraduate courses. Establishing a fair grading system is something necessary, though it might be hard to compare different approaches and different paths that each student would undertake.

References

Montessori, M.; and George, A.E. 1964. *The Montessori method.* New York: Schocken Books.

Dohrmann, K.R.; Nishida, T.K.; Gartner, A.; Lipsky, D.K.; and Grimm, K.J. 2007. High school outcomes for students in a public Montessori pro-

gram. *Journal of Research in Childhood Education* 22(2): 205–217.

AI in K-12 Education

John Lee (Antioch University)

I believe that we are fundamentally overlooking the development of important concepts in K-12 education. Both in non-arithmetic skill building as well as unique perspectives on ourselves.

One of the major over-arching themes in K-12 education is the development of humanity. Early cultures tell us about our society and the origins government. Zoology tells us not only about the animal kingdom, but also what it means to be human. Mathematics teaches us about fundamental truths and beauty. Astronomy teaches us our place in the cosmos and inspires us to reach beyond our own limitations.

AI can also teach us about what it means to be human. It can teach us what humanity looks like when taken to different extremes and thus develop within ourselves a deeper understanding of each other and our differences. It can easily demonstrate the truth and beauty of mathematics and how it can be used to develop models of knowledge and behavior. Each one of these models can then provide us with a unique perspective into our own cognition, psychology and the perspective of our existence.

A solid foundation in mathematics will start with movement, which will flow from real object manipulation to imagination to abstract cognition. This is introduced with early arithmetic. However, there is no similar early introduction of non-arithmetic cognition such as logic, search, iteration (folding), etc... that are vital for all kinds of engineering and programming. Such professions are shown to be deeply imaginative from mentally stepping through a program's execution to predicting the voltage levels across a circuit diagram. Early introduction of agent-based models through games and puzzles could provide this foundation as well as begin to introduce concepts for later exploration such as search, string-replacement-iteration, planning, machine learning, etc.

What this solid foundation of movement and imagination provides is a deeper understanding of and greater passion for mathematics. By the time we get to techniques such as multiple-column multiplication or long division we are beginning to learn procedures. This is what will make or break the love of mathematics. Those that truly learn what is behind the procedure and can see it in their imagination will do well, those that learn to blindly follow the procedure will not.

By the time we get to the upper grades, so much of engineering and sciences are taught procedurally. It is a crucial time to emphasize the importance of true-understanding, but class sizes, time constraints and material creep will make this difficult. How much easier would it be if there is a number of early grades experiences that begin to magically resonate with what is being taught.

I hope to explore the earliest introduction of core AI concepts in a concrete way to develop technical imagination skills, get us to think about how we think and finding new confidences in ourselves as we explore what it means to be human.

The Role of Ethics in AI Education

Joshua Eckroth (Stetson University)

In CS education, and AI education in particular, ethics is too often treated as a side concern, addressed in isolation from more typical topics. We view ethics as a cross-cutting concern that helps inform AI students, researchers, and practitioners how to be good scientists and engineers. Here, we examine five topics that are typically included in an AI curriculum and their respective ethical dimensions.

(1) Search and planning: AI systems that are deployed into real-world settings will be expected to perform accurately and reliably. Consider a search procedure, marketed as "Astar," that does not always find an optimal path due to a non-admissible heuristic. Or consider a planning system that does not account for the "frame problem," makes a wrong assumption about the state of the world, and fails to observe before acting. These examples illustrate unquantified risk resulting from inappropriate algorithmic decisions.

(2) Knowledge representation (KR) and rea-

soning: A KR schema is a surrogate for real-world entities (Davis, et al. 1993), and rarely attempts to model all of their complexities. For example, discretizing the range of human relationships into friends, married, or "it's complicated" introduces ethical questions about whether and what kind of inferences can be accurately drawn. Yet, high fidelity representations and inferential expediency remain in constant tension.

(3) Probabilistic reasoning: While probabilistic knowledge helps avoid making strict claims when knowledge is insufficient, probabilistic reasoning rarely yields certain inferences. Instead, some kind of decision theory must be consulted, which brings ethical questions about estimating risk and utility.

(4) Machine learning (ML): Learned models can be difficult to trust due to their complexity. In this sense, interpretable models like decision trees are less risky than less-interpretable models like neural networks. In either case, trust can be enhanced with hold-out and cross validation techniques. ML is more than "picking the technique that gives highest accuracy." We should know that the technique is best suited to the task at hand, and be able to justify that decision.

(5) Robotics: Once equipped with actuators, robots enter the ethical dimension. Failing to send a "stop motor" command due to software flaws may result in disastrous consequences. Machine/human control handoff (Klein et al. 2004), sometimes realized as a big red button, is a moment of vulnerability that can be mitigated with better status reporting and situation awareness. These issues go beyond typical robot building challenges.

We have shown that ethics should be addressed throughout the AI curriculum. The need for ethics arises from the need to be sure we are building systems that are appropriate for real-world situations and usable by people who depend on their accurate and reliable functioning.

References

Davis, R.; Shrobe, H.; and Szolovits, P. 1993. What is a knowledge representation? *AI Magazine* 14(1): 17–33.

Klein, G.; et al. 2004. Ten challenges for making automation a team player in joint human-agent activity. *IEEE Int Sys* 19(6): 91–95.

AI Education through Real-World Problems

Mark Crowley (University of Waterloo)

It is increasingly essential that practitioners of AI and ML focus on building verifiable tools with solution bounds or guaranteed optimality. One of my aims for AI/ML students is to give them the skills to build algorithms and analytical tools for providing verifiable guarantees and quality bounds on classification, prediction, and optimization problems.

The usual approach in a maturing field such as AI/ML would be to establish engineering standards for tools and methodologies that provide verifiable quality bounds and guarantees. Yet, the development of relevant tools are still an emerging research pursuit. Witness the extensive interest in the probability bounded results of Bayesian Optimization, the expanding application of PAC learning algorithms, or the wide usage of Gaussian processes to represent uncertainty and guide efficient sampling.

There is a growing application of AI/ML algorithms to safety critical domains such as automated driving, medical decision making and analysis and financial management. Also critical is the growth of computational sustainability: application of AI/ML methods to natural resource domains, wildlife management, energy management, socioeconomic planning and climate modelling. These domains all involve huge societal investments and impacts. Planning is often over a long horizon so what are acceptable risks and uncertainties in short term problems can expand over time into huge errors which undermine results.

Teaching students about these problems and the tools to address them will have an immediate impact on the world. In AI education we need to develop a new nucleus of an engineering discipline for AI/ML that provides students the framework to navigate the ever-expanding set of computational tools for solving complex problems.

This is an education ethics issue as well. If we are turning out students with the answers

to the world's problems, they need to know how to justify those answers in a rigorous way. This is often not the primary focus of AI/ML research or education. Many students who study AI/ML will continue straight on to industry rather than further research, so they will need to know the best algorithms to apply to different classification, prediction, optimization problems. However, to be AI engineers will in a way will require students to know more theory than a fully applied program which teaches use of existing methods. They need to know enough about the underlying probabilistic model, the sample complexity and the relationship of prior, latent, and observed variables in order to understand how reliable the results of their models are. Students also need a strong grounding in classical as well as Bayesian statistics so that they can make the right methodological choices for the given situation and do more than simply showing a histogram or ROC curve for their problem to justify their performance.

So, I feel the future of AI/ML education, especially at the undergraduate and master's level, is increasingly going to be focused on making AI into a true engineering discipline where requirements, guarantees, and design are as critical as reducing raw error rates of a classifier.

Making AI Concepts More Accessible

Richard G. Freedman (University of Massachusetts Amherst)

While it may be unreasonable to expect early undergraduate and secondary school students to code AI algorithms, it is possible for them to visualize and experience these algorithms firsthand. Developing an understanding of AI through these perspectives may even facilitate abstract thinking and problem solving when learning computer science and programming later. Although taught later in the CS curriculum after students are comfortable with computational thinking, many topics in AI can be explained conceptually using only high school mathematics. However, the manner in which these concepts are taught needs to be less traditional.

Based on the average student's present-day lifestyle involving personal mobile devices and almost limitless access to media, most students are used to constantly interacting with others and/or engaging in entertainment. This nearly contradicts the traditional lecture style for presenting material impersonally at the front of the room using chalkboards or slides. Instead, students today are accustomed to short spurts of watching and then lots of time doing, which goes hand-in-hand with some elements of team-based learning. In particular, an instructor should only briefly introduce a topic and related activity. Then, the students may explore the activity in groups in order to experience the concept on their own, interacting with each other to understand what happens. For example, A* search can be performed with a map and deck of cards; each card covers a city and students write the ruler distance (Euclidean heuristic) on each card as it is added to the frontier. The visited cities' cards are stacked in a deck to visualize the visited sequence.

By focusing on the algorithms' processes rather than the specific implementation, younger students without computational experience, higher-level mathematics, and programming skills can participate. The early focus of AI was to emulate human intelligence, and these students can relate to that by wondering, "how would I solve this problem?" These are questions they can discuss with each other and the instructor while performing the activities. In particular, the instructor can now make her time with students more personal by visiting groups to discuss and give tips based on their progress. Groups can also interact with each other afterwards to compare results.

Just as important as the interaction in the classroom, time outside of class can be vital to learning. Besides homework assignments that review concepts, students spend time on the internet watching videos and listening to music. Educational content can be provided in such entertaining forms. Alongside the classic television series *Bill Nye the Science Guy*, on-line streaming services such as YouTube have channels devoted to fun, short videos teaching mathematics (Vihart) and science (Veritasium, VSauce). While such a channel does not seem to exist for AI outside of Michael Littman's music videos, it is possible to present real-world examples and per-

form the activities above to create one. Then younger students are exposed to AI topics at any time in formats that they are more ready to digest, using high school-level knowledge without focusing on the code.

Rethinking the AI Ethics Education Context

Rogelio E. Cardona-Rivera (North Carolina State University)

Ethics, the moral principles that govern a person's or group's behavior, cannot be incorporated into a curriculum around AI without a systematic revision of the surrounding context within which AI takes place. We must go beyond just talking about ethics in the classroom; we need to put ethics into practice. I offer three recommendations for doing so, drawn from how ethics are treated within engineering and the social sciences.

Firstly, the Association for the Advancement of AI (AAAI) should institute an association-wide code of ethics. This recommendation is inspired by ethics codes in engineering, which include concern for the public good as a constituent part. For instance, the code of ethics of the National Society of Professional Engineers (2007) contains seven fundamental canons, the first of which is: "Engineers, in the fulfillment of their professional duties, shall hold paramount the safety, health and welfare of the public." An association-wide code of ethics would formally recognize our impact in and the responsibility that we owe to our society.

Secondly, research funding applications that deal with AI should be required to assess risks to society. This recommendation is inspired by similar requirements by Institutional Review Boards (IRB) within the social sciences (e.g., U.S. Dept. of Health and Human Services 2009). Whenever researchers conduct studies that deal with human participants, they are asked by an IRB to assess sources of potential risk; AI research applications should do the same. Importantly, these risk assessments should consider threats beyond immediate physical harm; e.g., the development of new analytical tools for understanding large amounts of data may inadvertently make it easier to reconstruct personally identifiable information, which constitutes a threat to anonymity, and which may disadvantage vulnerable populations.

Thirdly, students in AI project-based courses should be required, as part of the class' deliverables, to submit documents that assess the impact to society (in the context of the proposed AAAI code of ethics, and which should include an IRB-like risk assessment). Ideally, AAAI would serve as a facilitator of this kind of assessment, by providing a library of case studies and expert testimonies that can guide students in examining the broader implications of their work.

Incorporating ethics into a curriculum is more than a one-shot affair. It requires a systematic revision of the surrounding context within which AI exists, in terms of how we talk about it (first recommendation), how we fund efforts in it (second recommendation), and how it is put into practice (third recommendation). By leveraging existing models on ethics from engineering and the social sciences, we will be better equipped to offer concrete recommendations to ensure that ethics aren't an afterthought, but are integral to the development of AI.

References

National Society of Professional Engineers. 2007. *Code of ethics for engineers*. Technical Report 1102.

U.S. Department of Health and Human Services. 2009. 45 CFR 46.111: *Criteria for IRB approval of research*. Technical report.

Lifelong Kindergarten for AI

Tiago Machado (New York University)

To meet the expectations of young generations (who are highly exposed to games and other virtual interactions) regarding an introductory AI course, our purpose is to design a course based on the principles of the Lifelong Kindergarten (LK) (Resnick, M. 2007) and the Zone of Proximal Development (ZPD) as fields for dialogue (Meira and Lerman. 2001). From the former, we follow the principles of imagine, create, share and reflect. From the latter, we follow the idea of using it as a way to improve class communication with and among students.

We plan to create a 12-week course (2-hour class) in a level suitable for secondary school students with previous experience in programming languages. As video games are an attractive media to our audience, the course will use the General Video Game Framework (GVG-AI) (Perez-Liebana et al. 2016), which allows developers to implement algorithms to play famous arcade games. The course will have three stages: 1) Introduction to the GVG-AI, 2) Search Algorithms and 3) Supervised Learning Algorithms.

The first stage (Introduction to GVG-AI) explains how to work with the framework. It guides the students through a set of simple examples, followed by simple assignments, like creating an agent that plays the games by choosing random actions. The second and the third stages present the same structure: in the first week, the instructor explains the algorithms. Afterward, the students will have three weeks to implement the algorithm assigned to their group plus a class presentation. During these weeks the course will work in a blending class format. The students will have total access to videos, books, software, and the instructor to study and learn how to implement the algorithms in the GVG-AI framework.

It will be required that the presentation should not be a traditional one (i.e., students presenting slides and speaking about what they did). Fun and play with the content should be encouraged. As well as taking extra care about actually teaching to others how they can obtain the same results.

This way the students will be more active by imagining and creating their solutions. During the presentations, we will exercise more the share and reflect principles of the LK. The students will be encouraged to ask questions and share (all the resources they used to learn and implement, including the resulting code) their solutions with the class.

Throughout the course, instructors should be aware of students' progress. They should create social network or email channels to connect with students both in- and outside class. In this way we exercise, in both physical and virtual situations, the ZPD function of an intersubjective space via activities in which participants teach and learn from each other.

References

Meira, L.; and Lerman, S. 2001. *The zone of proximal development as a symbolic space.* South Bank University.

Perez-Liebana, D.; et al. 2016. The 2014 general video game playing competition. *IEEE Trans. Comp. Intel. and AI in Games* 8(3).

Resnick, M. 2007. All I really need to know (about creative thinking) I learned (by studying how children learn) in kindergarten. *ACM SIGCHI Conf. on Creativity & Cognition* (C&C '07).

Training Students in AI Ethics

Tom Williams (Tufts University)

After completing a course in AI, it is generally assumed that a student will be able to (1) characterize the task environment of a new problem, including the performance measure which should be optimized in that environment, and (2) identify the design tradeoffs between different algorithms for solving that problem. Unfortunately, students are rarely taught to consider the ethical facets of task environments that should be taken into account when deciding on performance measures and considering design tradeoffs, leading to blind spots for ethical failures in algorithm design.

In order to remove this blind spot, I believe that educators should strive to achieve the following learning objective: Students should be able to identify circumstances in which a tradeoff must be made with respect to task performance and ethical performance (especially with an eye towards verifiability), and be able to argue why a particular choice of algorithm strikes an appropriate balance between task performance and ethical performance.

One way to fulfill this objective could be to train students to evaluate proposed AI solutions by asking the following:

Consequentialism: (1) Is it possible that a decision made within this problem domain could harm another agent? (2) If so, can you guarantee that the proposed approach will find the solution that does the least harm (or harm below some justifiable threshold)? (3) If the answer to 2 is no, is there any other known AI solution for which the answer is yes? (4) If the answer to 3 is yes, what is the justification for the use of the proposed algorithm? If the

answer to 3 is no, what is the justification for solving this problem computationally?

Deontology: (1) Is it possible that a decision made within this problem domain could violate a legal statute or moral norm? (2) If so, can you guarantee that the proposed approach will find a solution that results in the fewest rule violations (or violations below some justifiable threshold)? (3–4) Same as above.

Virtue Ethics: (1) Is it possible that a decision made within this problem domain could be legal, and avoid explicit harm, yet fail to align with human virtues? (2) If so, can you guarantee that the proposed approach will find a solution that results in optimally virtuous behavior (or achieves a level of virtue that is above some justifiable threshold)? (3–4) Same as above.

The purpose of using this framework is to force students to "think like an ethicist" when designing or choosing between AI solutions: even though ethical concerns often present moral dilemmas to which there is no single obviously correct solution, students should get used to analyzing proposed solutions in order to identify possible ethical problems, identify what types of AI solutions make it difficult to verify or quantify ethical performance, and convincingly argue for or against a potential solution on ethical grounds.

Acknowledgements

The EAAI'17 New and Future AI Educator Program is partly supported by NSF Grant #1650295 and funding from the Artificial Intelligence Journal.

Eric Eaton is a faculty member at the University of Pennsylvania in the Department of Computer and Information Science, and in the General Robotics, Automation, Sensing, and Perception (GRASP) lab. His research is in machine learning and AI, with applications to robotics, sustainability, and medicine.

Sven Koenig is a professor in computer science at the University of Southern California. Most of his research centers around techniques for decision making (planning and learning) that enable single situated agents (such as robots) and teams of agents to act intelligently in their environments and exhibit goal-directed behavior in real-time.

Claudia Schulz is a postdoctoral researcher at the Ubiquitous Knowledge Processing (UKP) Lab at TU Darmstadt in Germany. She currently works on the automatic generation of feedback concerning the argumentative structure and content of students diagnostic essays, using various machine learning and NLP techniques. Her broader research interests also include knowledge representation and reasoning formalisms and their application for explanation.

Francesco Maurelli is Assistant Professor in Marine Systems and Robotics at Jacobs University Bremen, Germany, and Research Fellow at MIT, USA. His main research interests are in autonomy, in particular on active localisation, semantic understanding and fault management. He has participated and coordinated several international projects, with active cooperation with many universities and companies.

John Lee has recently turned his doctoral and industry experience in AI toward pedagogy, focusing on the elementary grades. He is primarily interested in teaching non-arithmetic mathematics and developing a strong foundation for technological imagination.

Joshua Eckroth is an assistant professor at Stetson University and chief architect of i2k Connect. His research interests lie in abductive reasoning and belief revision as well as computer science pedagogy. Eckroth holds a Ph.D. in computer science from the Ohio State University. He is coeditor of AITopics.

Mark Crowley is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Waterloo. His research seeks dependable ways to use Reinforcement Learning, Deep Learning and Ensemble Methods to augment human decision making when complexity arises due to spatial structure and uncertainty.

Richard Freedman is a Ph.D. Candidate in the College of Information and Computer Sciences at UMass Amherst. His research studies the integration of decision making and recognition for adaptive human-computer/robot interaction. He supports developing STEM activities for K-12 students and providing undergraduate students with opportunities to participate in research activities.

Rogelio E. Cardona-Rivera is an Assistant Professor of the School of Computing and the Entertainment Arts and Engineering Program at the University of Utah, where he directs the Laboratory for Quantitative Experience Design.

Tiago Machado is a PhD student in Computer Science at New York University. He obtained his B.Sc. and M.Sc. degree in Computer Science, both from the Federal University of Pernambuco. He worked as a STEM Consultant by using games to engage High School students. Nowadays, his main line of research resides in identifying how Artificial Intelligence can assist game designers.

Tom Williams is an assistant professor of computer science at Colorado School of Mines, where he directs the Mines Interactive Robotics Research (MIRROR) Lab. Toms research focuses on enabling and understanding natural language based human-robot interaction, especially as applied to assistive and search-and-rescue robotics. Tom earned a joint PhD in Computer Science and Cognitive Science from Tufts University in 2017.

AI Policy

Larry Medsker (George Washington University; lrm@gwu.edu)
DOI: 10.1145/3175502.3175510

Abstract

AI Policy is a regular column in AI Matters featuring summaries and commentary based on postings that appear twice a month in the AI Matters blog (https://sigai.acm.org/aimatters/blog/). Selected posts are summarized in issues of *AI Matters*.

Introduction

The SIGAI Public Policy goals are to:

- promote discussion of policies related to AI through posts in the AI Matters blog on the 1st and 15th of each month,
- help identify external groups with common interests in AI Public Policy,
- encourage SIGAI members to partner in policy initiatives with these organizations, and
- disseminate public policy ideas to the SIGAI membership through articles in the newsletter.

I welcome everyone to make blog comments so we can develop a rich knowledge base of information and ideas representing the SIGAI members.

Organizations Related to AI and Policy

In 2017 we expanded our reporting on and work with the American Association for the Advancement of Science, particularly the Center of Science, Policy, and Society. While AAAS policy issues are usually not directly related to AI, a regular look at their Policy Alert notifications is useful for larger policy issues, and helpful to see opportunities for SIGAI to be involved in public policy events. We also expanded our relationship with the ACM US Public Policy Council (USACM), which serves as the focal point for ACM's interaction with US government organizations, the computing community, and the US public in all matters of US public policy related to information technology. USACM addresses issues in innovation, privacy, security, digital governance, intellectual property, accessibility, and e-voting. I am a member of USACM, in part representing our ACM SIGAI.

Algorithmic Transparency and Accountability

ATA is a current major initiative with USACM. Your Public Policy Officer attended and reported on the USACM Panel on Algorithmic Transparency and Accountability, ehcih took place on Thursday, September 14th at the National Press Club. The panelists were moderator Simson Garfinkel, Jeanna Neefe Matthews, Nicholas Diakopoulos, Dan Rubins, Geoff Cohen, and Ansgar Koene. USACM Chair Stuart Shapiro opened the event, and Ben Sneiderman provided comments from the audience.

USACM and EUACM have identified and codified a set of principles intended to ensure fairness in this evolving policy and technology ecosystem. These were a focus of the panel discussion and are as follows: (1) awareness; (2) access and redress; (3) accountability; (4) explanation; (5) data provenance; (6) auditability; and (7) validation and testing. See also the full letter in the September 2017 issue of CACM.

The panel and audience discussion ranged from frameworks for evaluating algorithms and creating policy for fairness to examples of algorithmic abuse. Language for clear communication with the public and policymakers, as well as even scientists, was a concern—as has been discussed in our Public Policy blog. Algorithms in the strict sense may not always be the issue, but rather the data used to build and train a system, especially when the system is used for prediction and decision making. Much was said about the types of bias and unfairness that can be embedded in modern AI and machine learning systems.

The essence of the concerns includes the ability to explain how a system works, the need to develop models of algorithmic transparency, and how policy or an independent clearinghouse might identify fair and problematic algorithmic systems. Please read more about the panel discussion at https://www.acm.org/public-policy/algorithmic-panel and view the video at https://www.youtube.com/watch?v=DDW-nM8idgg&feature=youtu.be.

AAAI Fall 2017 Symposium Series

This year's Fall Symposium Series (November 9-11) https://aaai.org/Symposia/Fall/fss17symposia.php provided updates and insights on advances in research and technology, including resources for discussion of AI policy issues. The symposia addressed topics in human-robot interaction, cognitive assistance in government and public sectors, military applications, human-robot collaboration, and a standard model of the mind. Important themes for public policy were about the advances and questions on human-AI collaboration.

The cognitive assistance sessions this year focused on government and public sector applications, particularly autonomous systems, healthcare, and education. Human-technology collaboration advances involved discussions of issues relevant to public policy, including privacy and algorithmic transparency. The increasing mix of AI with humans in ubiquitous public and private systems was the subject of discussions about new technological developments and the need for understanding and anticipating challenges for communication and collaboration. Particular issues were on jobs and de-skilling of the workforce, credit and blame when AI applications work or fail, and the role of humans with autonomous systems.

IBM's Jim Spohrer made an outstanding presentation "A Look Toward the Future," incorporating his rich experience and current work on anticipated impacts of new technology. His slides are well worth studying, especially for the role of hardware in game-changing technologies with likely milestones every ten years through 2045. Radical developments in technology would challenge public policy in ways that are difficult to imagine, but current policymakers and the AI community need to try. See related references in the Resources section below.

Particular takeaways, and anticipated subjects for future blogs, are about the importance of likely far-reaching research and applications on public policy. The degree and nature of cognitive collaboration with machines, the future of jobs, new demands on educational systems as cognitive assistance becomes deep and pervasive, and the anticipated radical changes in AI capabilities put the challenges to public policy in a new perspective. AI researchers and developers need to partner with social scientists to anticipate communication and societal issues as human-machine collaboration accelerates, both in system development teams and in the new workforce.

Collaborations with other Policy and Ethics Groups

SIGAI recently began discussion with other groups, particularly between ACM and IEEE, on finding ways to bring together efforts on Algorithmic Transparency and Accountability. One opportunity is at RightsCon Toronto: May 16-18, 2018. The call for proposals mentioned "Artificial Intelligence, Automation, and Algorithmic Accountability' as one of their program "buckets."

As AI is becoming more pervasive in our lives, the impact on society is increasingly significant. Concerns and issues are being raised regarding value alignment, data bias and data policy, regulations, and workforce displacement. We need multi-disciplinary and multi-stakeholder efforts to find the best ways to address concerns using expertise from AI, computer science, ethics, philosophy, economics, sociology, psychology, law, history, and politics. AAAI and ACM are joining forces to start a new conference, the AAAI/ACM Conference on AI, Ethics, and Society. The first edition of this conference http://www.aies-conference.com will be co-located with AAAI-18 on February 2-3, 2018, in New Orleans. The program of the conference will include peer-reviewed paper presentations, invited talks, panels, and working sessions. The conference will cover a broad set of topics, including trust and explanations in AI systems, fairness and transparency, ethical design and development of AI

systems, and impact of AI on the workforce.

Upcoming

Some themes planned for the SIGAI Public Policy posts for 2018 include algorithmic accountability, human-machine collaboration, and the impacts of AI and Data Science on the future of education and the labor market. We will look at potential policies for today that could mitigate impacts of AI on individuals and society. Policy areas include innovative educational systems, ideas for alternate economic systems, and regulatory changes to promote safe and fair technological innovation.

Resources

- AAAI information related to science policy issues:
  https://aitopics.org/search
- AAAI Symposium Series:
  https://aaai.org/Symposia/Fall/fss17.php
- Ansgar Koene:
  https://theconversation.com/
  machine-gaydar-ai-is-reinforcing-stereotypes-that-liberal-societies-are-trying-to-get-rid-of-83837
- CACM letter on algorithmic transparency and accountability:
  https://cacm.acm.org/magazines/2017/9/
  220423-toward-algorithmic-transparency-and-accountability/
  fulltext#FNA
- Jim Sphorer, A Look Toward the Future:
  https://www.slideshare.net/spohrer
- Humans, robotics, and the future of manufacturing:
  https://www.engadget.com/2017/09/11/
  human-robot-ai-manufacturing/
- New education systems and the future of work:
  https://www.edweek.org/ew/articles/2017/09/
  27/the-future-of-work-is-uncertain-schools.
  html
- Noriko Arai's TED talk on Todai Robot:
  https://www.ted.com/talks/noriko_arai_can_a_
  robot_pass_a_university_entrance_exam
- Smart phone app "Seeing AI":
  https://www.youtube.com/watch?v=
  bqeQByqf_f8

Larry Medsker is Research Professor of Physics and Director of the Data Science graduate program at The George Washington University. Dr. Medsker is a former Dean of the Siena College School of Science, and Professor in Computer Science and in Physics, where he was a co-founder of the Siena Institute for Artificial Intelligence. His research and teaching continues at GW on the nature of humans and machines and the impacts of AI on society and policy[a]. Professor Medskers research in AI includes work on artificial neural networks and hybrid intelligent systems. He is the Public Policy Officer for the ACM SIGAI.

[a] http://www.humai.org/humai/ and http://humac-web.org/

Unemployment in the AI Age

Grace Su (Mounds View High School; graceduansu@gmail.com)
DOI: 10.1145/3175502.3175511

"Work saves us from three great evils: boredom, vice, and need" Voltaire said (Johnson & Indvik, 2004). Work is essential to the meaning of human life. To many people today, work means having a job. However, artificial intelligence (AI) will soon be able to take over many human jobs. A significant amount of social unrest will be caused by unemployment before the ethical issues of AI can be addressed. Thus, unemployment will be the most pressing social issue with respect to AI technologies.

AI has already surpassed certain human capabilities. In 1997, IBM's supercomputer Deep Blue defeated world chess champion Garry Kasparov (Newborn, 2012). In 2011, IBM's Watson stunned the technology industry with its victory against two of Jeopardy's greatest champions (Best, n.d.). In 2016, Google's DeepMind AlphaGo defeated the number one-ranked human Go player Lee Se-dol (Byford, 2016/n.d.). AI is no longer only a science fiction fantasy or a simple computer program that plays games; it has developed certain cognitive characters of the human brain that can learn and generate its own responses without explicit programming. Hence, with continued investment, AI will grow exponentially and dramatically transform society. Job automation with AI is becoming the prevailing trend across different industries. Some have even called AI the fourth industrial revolution, after steam power, electricity and electronics (Schwab, 2017). But, unlike past revolutions, this revolution could leave up to 35% of all workers in the UK, and 47% of those in the US, at risk of being displaced by technology over the next 20 years, according to Oxford University research (Stewart, 2015/n.d.).

In the past, many jobs have been lost due to technological advancements and large amounts of social upheaval were created. One classic example of this occurred during the 19th century's Industrial Revolution in England. As the use of automated looms and knitting frames increased, British weavers and textile workers who spent years training in their craft feared that less skilled workers were robbing them of their livelihood. However, the artisans' appeals for government assistance were ignored, so a few desperate weavers began breaking into factories and destroying textile machines. The people called themselves the Luddites and resistance against automated weaving spread across the English countryside. Resistance was so fierce that sledgehammer-wielding Luddites attacked and burned factories. . . in some cases they even exchanged gunfire with company guards and soldiers (Andrews, 2015/n.d.). The workers set upon these raids in hopes that the British government may ban weaving machines, but the government quashed the movement and made machine breaking punishable by death.

There also were many jobs lost between the 60s to 80s due to automation and outsourcing overseas for cheap labor. The consequences were significant. The Rust Belt in the US is the result of this mass job loss. In a discussion about his book "The New Minority: White Working Class Politics in an Age of Immigration and Inequality", Justin Gest described the pain and downward spiral of damage caused by job loss when many steel mills closed in Youngstown, Ohio. He stated that the city lost 50,000 jobs in about five years. During that time, suicide and divorce rates skyrocketed, and the city became the murder capital of the US by the late 80s. The city population dropped from 170,000 to about 65,000. As a result, many people felt they were marginalized and that they no longer had a voice in public policy, business interests and government (Gest, 2016). "If you look back to the first machine age the vast majority of Americans worked in agriculture. Now it's less than two percent," says economist Erik Brynjolfsson (Heath, n.d.). Today, the Luddites are simply remembered as technophobes, but they are a real example of the fear the threat of structural unemployment creates. And as demonstrated by Youngstown, that fear is not unfounded.

However, Brynjolfsson also states that "The computer processor doubles in power every 18 months, 10 times greater every five years, it's a very different scale of advancement and it's affecting a broader set of the economy than the steam engine did, in terms of all the cognitive tasks. It's happening a lot faster and more pervasively than before" (Heath, n.d.). Because of the swiftness of technology development, AI unemployment is a looming problem that needs to be confronted now.

In the near future, AI will cause unemployment as other technology did in the past, and dangerous social unrest will be provoked, perhaps at an even faster pace according to Brynjolfsson: "Unlike much of the 20th century we're now seeing a falling ratio of employment to population and that's something that concerns us. We don't think it's inevitable but we do think that many of the underlying trends in technology are likely to accelerate this so it's something we need to pay some serious attention to" (Heath, n.d.). Displaced workers are a political force that cannot be ignored, especially in elections. What they want are well-paid jobs that can support their families. Government, industries, and organizations should address unemployment issues caused by AI rather than brushing them off and claiming there will be more jobs created in the new technology revolution. Politicians in the past have promised to bring jobs lost due to outsourcing back to US. However, many traditional manufacturing jobs will never come back. According to research, while cheaper labor could cut cost by 60%, automation could cut labor costs by 90%. Even China, a country known for its abundant cheap labor and manufacturing jobs, has become a country with rising wages and labor shortages (Mahoney, n.d.). In 2017, China surpassed Japan as the number one country that uses the most industrial robots for manufacturing. Job automation is simply a natural choice for developed country to stay competitive in the global economy. With advances in automation and AI technology, some manufacturing factories will come back to US to get closer to the main markets. Unfortunately, moving back these factories will not create traditional manufacturing jobs; it will create more automated jobs instead. If unemployment continues, the income gap between the richest and the poorest people will get even wider. Income inequality will increase as owners of AI capital will expand their wealth while many workers may not receive benefits. Additionally, the wage gap between skilled and unskilled workers will increase. Long term unemployment cannot be risked as less income for families means low consumer demand that reduces income for businesses in other markets, gradually leading to a vicious cycle of economic downturn. A loss of dignity for workers may also accompany the unemployment AI will cause. Social unrest, crime, homicide, riots, race tension, and other current social problems could be exacerbated. However, the root cause of these problems can often be attributed to unemployment and lack of career opportunities. AI has the potential to cause many economic and social issues by automating many jobs and leaving millions of workers unemployed.

Companies have already begun to develop AI technology that will become the new engine of the old and continued process of job automation. Many countries have already used basic automation technology to replace labor. For example, industrial robots are widely used in automobile production lines across North America, Europe, Japan, South Korea, etc. However, with AI, job automation will be accelerated and it will have a broader impact. That is because traditional automation involves explicit programming, which requires a lot of human labor, and consequently increases production costs. In contrast, AI can utilize machine learning without explicit programming by evolving itself with the empirical data it collects from experimentation. For instance, Google's AI, AlphaGo, played Go against itself millions of times before competing with the best human player (Byford, 2016/n.d.). AI's capabilities can be improved dramatically and swiftly through repetitive training and around-the-clock experimentation through trial and error. Such a powerful learning system can be easily applied in many areas of life. One example of today's usage of machine-learning AI is the self-driving car (Giarratana, 2016). It would be extremely difficult and inefficient to develop self-driving technology through a traditional explicit programming method. It would require software engineers to account for all possible scenarios that could happen during real-life driving. Such conditions are

unlimited and are impossible to cover completely. But with AI technology, the car's system could make decisions that mimic a human driver's and learn from driving data shared from other self-driving cars. Google, Tesla, Uber, Ford, GM, BMW and most major automobile manufacturers are developing self-driving technologies. Many of them have already conducted hundreds of thousands of miles of real road testing. Such technology in certain aspects has surpassed human drivers because they have long-range radar, LIDAR, cameras, short/medium-range radar, ultrasound, and GPS, giving self-driving cars a multitude of senses that are superior to a human's (Giarratana, 2016). Self-driving cars can also communicate with each other wirelessly to know each other's driving intentions, make decisions and reactions much faster than a human, and do not get distracted or tired like a human. Therefore, self-driving cars are safer and could significantly diminish traffic fatalities. In fact, self-driving cars could produce a crash rate reduction of up to 90 percent (Ramsey, 2015). It is certain in the near future that many drivers' jobs will eventually be replaced by self-driving AI technology. However, today's progress in developing self-driving cars also demonstrates that AI has the potential to be ready to take over many other jobs very soon.

Even just a big loss of driving jobs will create a large impact on society. An estimated five million professional drivers' jobs, including truck drivers, taxi drivers, school bus drivers, and transit bus drivers, could be lost to self-driving technology (Greenhouse, 2016; Fahey, 2016). These drivers usually do not have higher education or other skills to find better replacement jobs. Their possible alternative jobs are limited and shrinking as the bar for education and skills to enter the future job market rises. Many families rely on the driver as the main income provider of the family. If 3/5 of the 5 million drivers have a family of four people, it would amount to about 12 million people's lives being affected by unemployment. Permanent displacement of these drivers would have devastating effects on families and their children. This negative impact on drivers' lives could ripple through multiple generations without external help. Additionally, many jobs areas supporting drivers such as driving schools, gas stations, and car dealerships may also disappear. For instance, lower accident rates would lead to less frequent visits to auto body repair shops, and that would leave a significant portion of the 445,000 auto body repairers in the US without a job (Lee, 2015). If losing jobs in one sector can cause this much damage, then society must prepare itself carefully for the unemployment AI could create.

Besides driving jobs, a variety of other labor positions could be replaced by AI automation. According to McKinsey & Company, five factors determine if a job can be automated: technical feasibility, cost to automate, relative scarcity, skills required, and cost of workers (Chui, Manyika, & Miremadi, 2015/n.d.). There are a wide range of job areas that are subject to AI takeover, such as transportation and logistics, office and administrative support, personal and domestic service, accounting, and construction (Rotman, 2013). Production lines that manufacture large quantities of fixed design are also candidates because they usually contain jobs that are repetitive. Many of these jobs are already automated or in the process being automated with consideration of the 5 factors mentioned above. Additionally, AI technology in image and sound recognition has surpassed humans, so many security jobs will be replaced by AI. Also, there are already robots that can flip burgers. Food preparation could be replaced by AI that is capable of controlling multiple factors precisely, including temperature, cooking time, color, smell and taste. In construction, there are already bricklayer robots that are in use. A company called Apis Cor used a 3D printing robot and built a house in Moscow in 24 hours (Cor, 2017). The house can last up to 175 years and costs only $10,134 to build. With AI technology, a machine may be able to understand a building plan without requiring a human programmer to program the building plan into the system. Warehouse workers can be replaced by AI robots that know the precise location of stocked goods and executes orders tirelessly and efficiently. In farming, AI technology could also help farmer detect plant diseases earlier to take preventive actions. It is possible that AI will become a powerful general-purpose tool that is used every day for work and study like today's personal

computers and smart-phones. Therefore, in low or medium skilled labor, AI automation will push wages lower than ever, making the less-educated work force even poorer.

Unlike previous industrial revolutions, AI could automate a multitude of jobs throughout the occupation spectrum, including intellectual jobs. But, there are not enough jobs that could quickly absorb displaced workers without extensive retraining. Breakthroughs in AI-based voice recognition have made automated voice search and personal assistants practical to use. Siri, Cortana, Alexa and Google Now are current examples of such systems. When these technologies mature, jobs such as personal assistances and foreign language translators will be eventually replaced by AI. Traditional high-paying jobs such as Wall Street traders could also lose their jobs. For example, Goldman Sachs used AI to replace 600 traders and left only two traders remaining, who were supported with 200 computer engineers (Byrnes, 2017). IBM's Watson AI system has already replaced 34 workers in a Japanese insurance company called Fukoku Mutual Life Insurance to calculate payouts to policyholders. The system is based on IBM's Watson Explorer, which possesses cognitive technology that can think like a human, enabling it to analyze and interpret all of your data, including unstructured text, images, audio and video according to the company (Asia, 2017). While IBM's Watson system has more knowledge than most humans, even its abilities to use knowledge is no match for a human's yet; however, it is powerful enough to automate a lot intellectual work to help human to make decisions. Some routine data collection and processing jobs such as legal assistant and accounting clerk are also subject to replacement by AI. Many other traditional white-collar jobs such as financial advisor, bank teller, and insurance underwriter could be largely automated or replaced by AI. As a result, it is not just the low wage, less educated work force, that will be affected, as many careers that require higher education could also be eliminated. According to 2013 government statistics, the US has 36 million (27.2% employment share) jobs that require less than high school diploma, and 51 million (39% employment share) that require a high school diploma (Williams-Grut, 2016). In total, 87 million (66% employment share) jobs require a high school education level or higher, making it difficult for workers to climb up the employment ladder even if AI does create new job opportunities while replacing traditional jobs. In conclusion, the impact of AI-caused unemployment should not be underestimated.

To confront this issue and reduce unemployment caused by AI, it is largely a matter of looking at the issue economically and adjusting the supply and demand of labor so that employment levels will have room to grow. Demand for labor comes from employers and is derived from consumer demand for goods. Supply for labor comes from how many workers are willing and able to work in the job market. Jobs that can be replaced by AI will have their demand of labor lowered due to improvements in capital, causing unemployment. Organizations and governments could find ways to increase demand of labor in other job areas to prevent structural unemployment from becoming permanent. On the demand side, common economic policies to raise employment includes injecting more money into the economy to increase consumer demand and therefore labor demand, giving government subsidies to companies that hire the long-term unemployed, and lowering employment taxes. On the supply side, employment can be raised through increasing investment in human capital to lessen occupational immobility and decreasing housing costs to reduce geographical immobility (Pettinger, 2017). These common policies and similar ones can be implemented by governments, industries, and organizations to prepare for the new wave of unemployment AI will create. Organizations can also discuss how to encourage the economy to allow job automation to lead to more profit, consumer demand, and therefore labor demand, instead of letting AI cut work too drastically. The funds for these methods can be raised through private sector contributions, taxes, or donations. These standard economic policies' applications, if considered within the specific context of the AI industry, will then help prevent AI-induced unemployment.

Governments, industries, and organizations can furthermore discuss plans to expand new careers that will be created by AI to minimize

the impact of mass unemployment. For example, with more AI used in our society, there will be more demand for AI jobs and trainers of all kinds. AI safety engineers, maintenance engineers, machine designers, and ethics researchers are some examples of possible new jobs. Specialized skills such as data analysis will be needed to identify areas that can use data to train AI and prepare data to conduct the training. As more and more opportunities for AI applications appear, new jobs that are difficult to predict today will also be created. Humans will be able to focus on doing more rewarding work such as scientific discovery, problem solving, and innovative design. Jobs indirectly related to AI will also expand, such as educational content creation for people who will go into AI careers. Experts in careers that will benefit from AI technology will also be in demand as AI implementation needs deep knowledge of each field for successful specialization. For instance, a field as complicated as neuroscience will require experts to create a usable AI system, because AI systems are only as good as the people who design them. The decline of monotonous jobs will also allow for creative jobs to expand as people have more time to pursue less technical interests. However, Moravec's paradox predicts that manual jobs are sometimes more resilient to job loss than middle-skilled jobs (Heath, n.d.). As a result, job expansion may occur in low and high skilled jobs, but not in middle skilled jobs. Society must invest more in higher education to adjust to this new set of jobs.

9.5 million job openings are predicted to be generated by AI; however, most of these are only open to the well-educated (Williams-Grut, 2016). To solve the unemployment problem, education that teaches skills for AI professions will be the key. Similar to how in the 1950s, funding of secondary learning helped curtail the unemployment effects of the Industrial Revolution, funding of college-level learning will mitigate the negative effects of AI on the economy (Heath, n.d.). On the supply side, young people coming to join the workforce must be prepared through updated education. AI will likely become an essential tool that people use every day to work and study, like how computers are utilized today. Because how well people can work with AI will become an important skill that gets people hired, schools need to teach students to understand how AI works and how humans and AI could complement to each other. In order to succeed in the future, young people should at least graduate from high school. Every high school and college could teach machine learning AI just like how people today are being taught computer skills and programming in school. Education will help develop people's problem solving and creative thinking skills. The education system should prepare the work force's critical thinking skills, skills to apply knowledge, skills to solve hard problems, and effective communication skills. These abilities are essential to adapt to the new economy supported by AI technology. Current college education can be transformed from focusing on transferring knowledge to train students to apply knowledge, solve hard problems. Student performance should not measure how well a student memorizes facts but how well they apply the knowledge they have learned. Training for jobs complementary to AI instead of competitive with AI should be promoted. Since most AI-related jobs lie in STEM areas, and demand for STEM workers will increase before the number of young people going into STEM jobs increases, governments, industries, and organizations should continue to encourage young people to pursue STEM careers. They could provide scholarships, grants, low-interest loans, or even waive the repayment of loans if students can graduate with a 4-year STEM degree from a public university. Education must be reformed to emphasize teaching students how to think and solve difficult problems so they can succeed as workers in the AI age.

For unemployed adult workers, governments, industries, and organizations can use AI to analyze job opening data and provide targeted, customized training in specifically demanded areas to the unemployed. AI can be used to create online skill profiles for every worker and then match workers with job openings across the country. In the current job market, there are job openings, but people who need jobs often do not find the best job for themselves. If AI could do the hard work to find a job that matches best for each person, it will boost the job market and bring supply and demand together more efficiently. Government programs

that provide funding for job training programs and unemployment support such as the US Department of Labor's Employment and Training Administration could also be considered in plans to increase employment. Loans provided by these programs can be earned back by the unemployed through paid work during training. The pay could include cost of training, basic food, housing, and child care if needed, so the unemployed can have the time and resources to go through the required training. Additionally, basic income and public works projects are also to be considered in preventing job loss by AI. For people who still have jobs, lifelong learning is needed to keep up with the advancements of technology. AI not only drives the change of education; it also provides a tool to adjust to such a change. Society should encourage more companies to provide high quality free online learning using AI technology to support adaptive learning. Some free online learning providers such as Khan Academy have already started experimenting with adaptive learning. That means AI could be used to understand how individual learns more effectively than the individual himself. Hence, these organizations will be able to provide high-quality, free public education and deserve direct funding from the government.

Education alone will not solve the unemployment problem if the long run goal to develop AI is to cut labor costs. Society can learn how to integrate AI and people together and produce even more powerful results. Humans and AI have been proven to achieve more than using just one or the other (McCorduck, 2004). AI is ideal for specialization, thus it can be a great tool for humans instead of a complete replacement for humans. As a result, AI will likely replace tasks rather than complete jobs. Humans will still cover jobs in areas where the morals required for tasks are not present in AI, as Joseph Weizenbaum says (McCorduck, 2004). This can allow for a more gradual and society-friendly transition into the AI age. Additionally, AI can allow people to flow between jobs with reserves of corporate knowledge available at all times for humans to apply. Ultimately, AI are machines created by humans and should be used as a tool by humans. Governments and companies will use AI to help to make strategic decision for policy changes, investment or marketing or estimate risks. More and more intellectual work will require human to work with AI to form a team. For example, with AI diagnostic assistance, doctor could spend less time reading diagnostic images and data, spend more time with the patient, and refer to AI suggestions of diagnosis for disease and treatment to serve the patient better. Therefore, how well a person work with AI will become an important determinant of one's employability. It is likely that only certain skills would need to be taught and a lot more knowledge could be accessible by people. Then, people would have more time to create and human creativity could expand. There will not be many jobs that can progress without incorporating AI technology.

Providing the unemployed affordable health care is also important to preventing social unrest. Current living and health care costs are not affordable for most of the unemployed. Besides providing education and retraining, society should also take advantage of AI technology to drive down living and health care costs. This will help let the benefits of AI be shared with everyone. Without control of health costs, a social safety net that includes retraining for the unemployed would not be sustainable. The US spent 17.1% of its GDP on health care in 2013, yet health costs are still rising quickly today (Squires & Andersons, 2015). Fortunately, AI has huge potential to drive down the health cost and make health care, especially preventive care, available to most people. For example, there are studies showing that AI can be trained to detect skin cancer more accurately than many expert human dermatologists (Scutti, 2017). Additionally, IBM is partnering with Celgene to better track negative drug side effects (Leaf, 2016). They are applying IBM's cognitive computing AI technology to recommend cancer treatment in rural areas in the U.S., India, and China, where there is a dearth of oncologists. In fact, IBM's Doctor Watson systems are designed to become a medical expertise system, and it can complete high-expertise tasks such as initial diagnoses of illness. The computer system can store vast amounts of information about illnesses and analyze a patient's symptoms to provide an initial diagnosis or what test to do to further collect information. Such a system could even be sold on phones or in stores in the future, with abilities to take blood pressure,

heart rate, blood samples, or urine samples to drive down health care costs and make health care affordable. For health care, early detection and action are crucial to lowering expenses. Decreasing living expenses through AI will then allow unemployed workers to invest more money in retraining programs.

Another danger to employment is the barriers to entering the AI technology industry. A concentrated AI market will stunt job growth and fuel unemployment instead because innovation will be limited. It is a fact that small businesses create the most jobs in America and in many other countries in the world. According to U.S. Small Business Administration, small businesses create 64% of private sector jobs (*Frequently Asked Questions about Small Business*, 2012). Currently, the AI industry has some high barriers, including the requirement of a massive amount of data, a vast amount of storage space, and immense computing power to enter the business, and only a small number of companies can afford the research facilities and resources required for AI development. Society should not let these barriers get so high that only a small number of companies that have the most data can develop advanced AI systems. If left unchecked, these companies would be able to dominate the AI market with their influence. They could also control and dominate the government and military and gain considerable political power. This limitation of access to data stunts industry growth and inhibits innovation, as concentrated markets often do. Data is the most important resource in the AI age. There are currently a lot of opportunities in society to apply AI. A few big companies will not be able to cover all of them efficiently. The government needs to provide more support to give any person the ability to use AI to create new technology by making some of the data the government owns free to the public, and providing storage and computing power to develop AI. Then, government data on subjects such as population, traffic, weather, satellite images, and the human genome could be used to create many jobs using AI. This action will also encourage more people to work and invest in AI. AI could help us perform scientific discovery more efficiently, find a cure for cancer, devise a solution to global warming, conceive a way to conserve nonrenewable resources, or discover other solutions to global issues. Thus, the government should consider restricting technology giants such as Google, Amazon, and Apple from buying up the majority of AI startups. Additionally, rapid development is occurring in AI without public discussion or supervision of ethics because there are only a few large businesses involved. For example, “Google’s ethics board is shrouded in secrecy, with both Deepmind and Google refusing to disclose any details about the members of the board or what is discussed” (Wales, 2016). So it is important to limit the influence of business by maintaining competition in the AI market. AI, a technology with such powerful potential, should not be entrusted to the hands of a few private corporations and thus the public must keep watch on it and prevent monopoly.

In summary, AI will accelerate job automation in the world and cause structural unemployment. There are opportunities and challenges with such a change. Technology revolutions have already caused unemployment before, with serious economic and social consequences. Governments, industries, and organizations can discuss how to encourage expansion of jobs created by AI to replace jobs destroyed by AI. Strategies to diminish unemployment include reforming education, boosting federal programs that provide training and support to workers, combining AI and humans to perform jobs, driving down health and living costs, and keeping the AI industry competitive. Automation may become so ubiquitous that displaced workers will have nowhere to go with their current skills, producing social upheaval. Society cannot afford to assume that AI will create jobs as fast as it will eliminate jobs or distribute new wealth to all people like other technology has before. To quote from Brynjolfsson: ”It’s one of the dirty secrets of economics: technology progress does grow the economy and create wealth, but there is no economic law that says everyone will benefit” (Rotman, 2013). Therefore, we must act now in order to prepare our current and future workforce for widespread use of AI and spread the bountiful benefits of automation out so everyone can have a future of increased leisure, wealth, and freedom.

References

Andrews, E. (n.d.). *Who were the luddites?* Retrieved February 27, 2017, from http://www.history.com/news/ask-history/who-were-the-luddites

Asia, . B. N. (2017). *Japanese insurance firm replaces 34 staff with ai.* Retrieved February 27, 2017, from http://www.bbc.com/news/world-asia-38521403

Best, J. (n.d.). *Ibm watson.* Retrieved February 27, 2017, from http://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/

Byford, S. (n.d.). *Google's deepmind defeats legendary go player.* Retrieved February 27, 2017, from https://www.theverge.com/2016/3/9/11184362/google-alphago-go-deepmind-result

Byrnes, N. (2017). *As goldman embraces automation, even the masters of the universe are threatened.* Retrieved February 27, 2017, from https://www.technologyreview.com/s/603431/as-goldman-embraces-automation-even-the-masters-of-the-universe-are-threatened/

Chui, M., Manyika, J., & Miremadi, M. (n.d.). *Where machines could replace humansand where they cant (yet).* Retrieved February 27, 2017, from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet

Cor, A. (2017). *The first on-site house has been printed inrussia.* Retrieved February 27, 2017, from http://apis-cor.com/en/about/news/first-house

Fahey, M. (2016). *Driverless cars will kill the most jobs in select us states.* Retrieved February 27, 2017, from https://www.cnbc.com/2016/09/02/driverless-cars-will-kill-the-most-jobs-in-select-us-states.html

Frequently asked questions about small business. (2012, September). Retrieved February 27, 2017, from https://www.sba.gov/sites/default/files/FAQ_Sept_2012.pdf

Gest, J. (2016). *The new minority: White working class politics in an age of immigration and inequality*. Oxford University Press.

Giarratana, C. (2016). *How ai is driving the future of autonomous cars.* Retrieved February 27, 2017, from http://readwrite.com/2016/12/20/ai-driving-future-autonomous-cars-tl4/

Greenhouse, S. (2016). *Autonomous vehicles could cost america 5 million jobs. what should we do about it?* Retrieved February 27, 2017, from http://www.latimes.com/opinion/op-ed/la-oe-greenhouse-driverless-job-loss-20160922-snap-story.html

Heath, N. (n.d.). *Why ai could destroy more jobs than it creates, and how to save them.* Retrieved February 27, 2017, from http://www.techrepublic.com/article/ai-is-destroying-more-jobs-than-it-creates-what-it-means-and-how-we-can-stop-it/

Johnson, P. R., & Indvik, J. (2004). Digital depression, stress, and burnout. In *Allied academies international conference. academy of organizational culture, communications and conflict. proceedings* (Vol. 9, p. 19).

Leaf, C. (2016, November 01). *Heres the surprising reason ibm is partnering with celgene.* Retrieved February 27, 2017, from http://time.com/4552874/ibm-watson-health-celgene-partnership/

Lee, J. (2015). *Self driving cars endanger millions of american jobs (and thats okay).* Retrieved February 27, 2017, from http://www.makeuseof.com/tag/self-driving-cars-endanger-millions-american-jobs-thats-okay/

Mahoney, S. (n.d.). *The real cause and impact of chinas labor shortage.*

Retrieved February 27, 2017, from http://www.manzellareport.com/index.php/special/the-real-cause-and-impact-of-china-s-labor-shortage

McCorduck, P. (2004). *Machines who think : a personal inquiry into the history and prospects of artificial intelligence* (2nd ed.). Natick, MA: A K Peters, Ltd.

Newborn, M. (2012). *Kasparov versus deep blue*. Springer Science & Business Media.

Pettinger, T. (2017). *Policies for reducing unemployment.* Retrieved February 27, 2017, from http://www.economicshelp.org/blog/3881/economics/policies-for-reducing-unemployment/

Ramsey, M. (2015). *Self-driving cars could cut down on accidents, study says.report predicts mass adoption of auto-piloted vehicles beginning in about 15 years.* Retrieved from https://www.wsj.com/articles/self-driving-cars-could-cut-down-on-accidents-study-says-1425567905

Rotman, D. (2013). *How technology is destroying jobs.* Retrieved February 27, 2017, from https://www.technologyreview.com/s/515926/how-technology-is-destroying-jobs/

Schwab, K. (2017). *The fourth industrial revolution*. Crown Business.

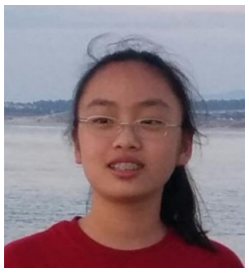
Scutti, S. (2017, January 26). *Automated dermatologist detects skin cancer with expert accuracy.* Retrieved February 27, 2017, from http://www.cnn.com/2017/01/26/health/ai-system-detects-skin-cancer-study/

Squires, D., & Andersons, C. (2015, October). *U.s. health care from a global perspective: Spending, use of services, prices, and health in 13 countries.* Retrieved February 27, 2017, from http://www.commonwealthfund.org/~/media/files/publications/issue-brief/2015/oct/1819_squires_us_hlt_care_global_perspective_oecd_intl_brief_v3.pdf

Stewart, H. (n.d.). *Almost half of all us workers are at risk of losing their jobs to robots.* Retrieved February 27, 2017, from http://www.businessinsider.com/almost-half-of-all-us-workers-could-lose-their-jobs-to-robots-2015-11

Wales, J. (2016, October 11). *Tech giants' artificial intelligence monopoly possibly the most dangerous in history.* Retrieved February 27, 2017, from http://theantimedia.org/tech-giants-ai-dangerous-monopoly/

Williams-Grut, O. (2016, 15). *Robots will steal your job: How ai could increase unemployment and inequality.* Retrieved February 27, 2017, from http://www.businessinsider.com/robots-will-steal-your-job-citi-ai-increase-unemployment-inequality-2016-2



Grace Su Grace is a junior at Mounds View High School in Arden Hills, MN. Grace loves programming and taught herself HTML, CSS, JavaScript, and Java in middle school. She is very interested in artificial intelligence's progress and its impact on human society. She wants to become an AI researcher after college.

You, Me, or Us: Balancing Individuals' and Societies' Moral Needs and Desires in Autonomous Systems

Joseph A. Blass (Northwestern University; joeblass@u.northwestern.edu)
DOI: 10.1145/3175502.3175512

Abstract

Autonomous systems are increasingly taking actions with moral consequences. While certain universal norms may be built into such systems, other areas should be personalizable to the end user. This essay motivates this problem, explores the ways such systems are already making decisions with moral ramifications, and proposes a path forward.

Introduction

We are on the cusp of a new era, in which autonomous systems make decisions with moral consequences that impinge on numerous aspects of our lives. AI programs ought to consider the moral ramifications of their actions (Scheutz, Malle, & Briggs, 2015), but can designers engineer for such systems a single universally standard set of morals to cover all situations? It's unlikely: while the law takes a stand on certain matters, others are left to individuals, whose ethical and moral norms vary within and across societies. Any complete set of morals accepted by one person will somehow be in violation of another's. We equally cannot avoid building any ethical or moral reasoning capabilities into autonomous systems, or let them do whatever humans tell them: such systems could do great harm, either through ignorance or by human instruction. The core ethical challenge facing AI researchers and engineers is balancing individual ethical and moral preferences with the norms needed for society to function.

This essay will argue why there are many situations in which individuals should have some control over their machines' ethics. It will argue that the solution to this problem is to collectively agree upon a set of universal principles for autonomous systems to abide by, but which will cover *only a portion* of all possible ethical and moral situations, with users having some say over the rest. These built-in norms will be those which are core to society and necessary for humans to trust these systems. Building such a system will be time-consuming. It will involve philosophers, ethicists, sociologists and psychologists working to determine our minimal collective norms; AI experts to implement them; and lawyers and lawmakers to write the laws regulating them.

Morals Vary Within & Across Societies

Morals are principles cleaving right from wrong, and ethics are codes of conduct based on morals. While these are fundamentally different, this essay treats these terms as interchangeable as they relate to AI. Human morality is partly rooted in emotion and instinct (Haidt, 2001); fully modeling it may prove to be an AI-complete problem. Constructing ethical norms that lead to decisions humans consider moral, however, is within the purview of current AI technologies.

While humans may generally agree on some moral norms (e.g., don't steal from or attack other members of society), there is no one universal moral system that we all subscribe to. The specific morals and ethics we develop are defined in large part by culture and experience. One well-studied example distinguishes between cultures that value individualism vs. collectivism. Broadly speaking, Individualist (generally Western) societies focus on individual actions and value personal freedoms, whereas Collectivist (generally Eastern) societies focus on social outcomes and value social harmony. Societies on different sides of this fault line show different patterns of moral judgment. In one case, researchers asked Canadian and Chinese children to evaluate lies that either helped or hurt a social group (Lee, Cameron, Xu, Fu, & Board, 1997). Chinese children judged prosocial lies more positively than antisocial ones, but Canadian children did not. For the Chinese children the relevant feature was the effect of the action on the group, whereas for the Canadian chil-

dren it was the nature of the action itself. Cultural moral differences like these are not easily dismissed, and it would be wrong for engineers from one culture to force members of another to use machines that abide by the culturally-specific ethical conventions of the first. (For a review of the role of culture in moral judgments, see (Sachdeva, Singh, & Medin, 2011)).

Even *within* societies, people vary dramatically in their moral judgments. Psychologists can identify overall trends in patterns of responses, but rarely find universal agreement. For example, Haidt and colleagues (Haidt, Koller, & Dias, 1993) asked participants in the US and Brazil whether violations of moral norms around purity (e.g., using a national flag to clean a toilet) were permissible. Most people in both countries identified these actions as moral violations, but rated them as being harmless. However, they disagreed about whether these actions should be permitted, and socio-economic status predicted their responses more than nationality. If people disagree within a society about whether certain types of harmless actions are permissible, it is hard to imagine a standards board determining a complete set of principles that all users will perceive as morally just without being overly restrictive or permissive.

Some may say harmless actions should not be regulated, but an AI system acting on a person's behalf should not take actions that are offensive, even if they cause no concrete harm. Robots shouldn't use flags as rags. Moreover, people disagree over how wrong certain unambiguously harmful actions are. In the classic trolley problem, a trolley will hit five people, who can be saved by sacrificing one person. In the *switch* scenario, the trolley can be turned onto a side track with one person (who will die). In the *footbridge* scenario, a person can be pushed in front of the trolley, stopping it. People overwhelmingly say taking action in the switch scenario is better than taking action in the footbridge scenario, but some think acting in the footbridge scenario is permissible. Fully 60% of participants in a recent study endorsed acting in the footbridge scenario (compared to 79% for the switch scenario) (Hristova & Grinberg, 2016). Response patterns change if a humanoid robot is acting (the switch scenario is more permissible) and change again when a non-humanoid automated system takes action (both scenarios are more permissible). Some people see actions as more or less permissible depending on the identity of the victims (Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009). Clearly the reliable trends these studies reveal do not point to a universal "right" or "wrong" answer to these scenarios.

Certainly a society can forbid autonomous machines from taking certain types of action. Prohibiting unprovoked aggression and violence, theft, and abuse are obvious restrictions to place on computers. We have standards of psychopathy among other mental illnesses, and we should not build AI systems that meet their diagnostic criteria. Furthermore, human history and societies are filled with instances where powerful subpopulations claimed a moral right, if not imperative, to oppress and exploit other subpopulations. AI systems should never be allowed to promote oppression.

AI should also not be used to enable human hypocrisy. For example, Bonnefon and colleagues found that most people say autonomous vehicles should minimize loss of life on roads, even if doing so involves sacrificing the driver. These same people indicated that they would not want to buy such a self-driving car (Bonnefon, Shariff, & Rahwan, 2016). They are happy to impose on others an ethical system that they themselves refuse to follow. Nevertheless, we recognize that this desire cannot be satisfied in our interactions with other humans: we don't expect our doctor to sacrifice her other patients to save us. When deciding what norms to build into an autonomous intelligent system, people similarly must recognize that those ethical norms may limit the actions a machine can take on their behalf.

Nevertheless, there is a wide range of circumstances in which people should have some say over the behavior of their machines. These circumstances arise outside of those situations involving the core moral norms that protect society, and concern both positive and negative situations: not only those where someone is likely to get hurt, but also ones where the system will have to decide who, and how, to help. While this may sound like the

stuff of science-fiction, AI systems are already making these kinds of decisions, and it is problematic that they are not doing so with an explicit and consistent set of moral principles.

Moral Hazards for Autonomous Systems

Much ink has been spilled over the ethics of self-driving cars, but many of the trickiest and most difficult ethical questions arise with other, more mundane systems that are already in widespread use and affect our lives. One such class of system is machine-learning based law enforcement assistants. These are being used to predict flight and recidivism risk when determining bail at trial, and for facial recognition purposes for law enforcement. Since machine learning algorithms learn from their inputs, if the input encodes systemic bias, the algorithms will too. One recidivism predictor gives harsher scores to African-American defendants than to White ones, despite race not being an input criterion (Angwin, Larson, Mattu, & Kirchner, 2016). Criminal facial-recognition software disproportionately mistakenly identifies African-Americans, who are more likely to have had contact with the police, regardless of criminality (Garvie, 2016). Being mistakenly called in for questioning or unfairly denied bail can have a devastating effect on a person's life.

Reducing the bias within those systems is difficult (Diakopoulos, 2016), but if bias cannot be eliminated, the systems should not be used. If bias is eliminated, societies still must decide how confident their system's judgments must be. A system that requires a lower confidence threshold will miss fewer criminals but flag more innocents; a system with a higher threshold will do the reverse. Should the system be more concerned with security or liberty? This is a choice that must be made *by* the community, not *for* the community by the software company.

Or consider automated hospital management systems. These systems, which are in increasingly widespread use, perform a variety of operations including bed assignment and supply management. Both of these tasks have potential moral implications. In the midst of an epidemic, a hospital may run out of beds. On what basis will an automated system determine who gets a bed and who does not? How quickly (and which) patients that have beds will be discharged to make room for others? These decisions vary by hospital, by patient, and by epidemic. Doctors of course have the final say, but a well-designed automated system ought to facilitate that decision-making process. For doctors and administrators to rely on the system during a hectic situation, they must trust it to make the right decisions. This trust requires transparency and the knowledge that the system will implement the hospital's priorities.

Stocking medicines costs money. The factors that determine how much medicine of any kind to stock are particular to individual hospitals. What happens in the rare instances when they have more patients than medicine? Though unlikely, these situations are nearly guaranteed to occasionally occur, for example in the early stages of an epidemic. In such cases, which patient gets the medicine, and who must wait or get another treatment? Even using a first-come-first-serve basis is making an ethical choice. A patient might make a different decision about which hospital to go to if they think they are more likely to get a rare medicine at a hospital with a different disbursal process. Again, hospital administrators must understand and control how their systems make these decisions.

Automated hospital systems solve more problems than they create. They are more efficient and accurate, integrate more data, and are less biased than their human counterparts (i.e. they don't care whether a patient is rude). The point is not that they will be immoral or unethical, but that hospitals deal with ethical grey areas and develop their own ethical standards within clearly defined professional ethical standards. Decisions made by automated systems should be consistent with hospitals' developed norms.

This example also shows that a decision about who to hurt can be a decision about who to help. It is a truism that we cannot help everybody. A self-driving car may do the most good by abandoning its owner and driving to the country with the highest rate of malnourishment to volunteer itself for Meals on Wheels, but few would argue we should build such cars. Resources are limited, and AI systems

need to know who to help, and how. They must express positive values, not only avoid negative ones.

There is at least one product on the market which people will soon want to have express positive moral and ethical values, including ones they should have control over: Mattel's Hello Barbie. Hello Barbie is a doll with AI-driven conversational abilities that learns about your child. Hello Barbie received negative attention over security issues (it requires an internet connection and does all its computation on the cloud), but a Hello Barbie of the near future may run locally and avoid these concerns. Nonetheless even without any security concerns, such a system potentially involves a range of moral concerns, both positive and negative.

The largest of these is the moral development of the child. Children are constantly learning, and they learn how to be members of society through social interaction. We don't know how interacting with such a doll will affect a child's social and moral development, but children may well learn from their interactions with it. Whatever values the doll has been instilled with may potentially be learned by the child. Again, the doll should discourage harm, theft, etc. But what about the benevolent lying example discussed above? This (among others) would presumably be a value the child's parents would want to teach them. If the parents are teaching the child that otherwise harmless prosocial lies that promote social cohesion are OK, but the doll teaches the child that lying is never OK (even to protect someone's feelings), then the toymaker is directly undermining the parents' moral instruction. And what about religion? A parent raising a child in one religion might object to the doll expressing beliefs in another religion (or none), and a parent in an atheist household might object to the doll expressing any religious beliefs whatsoever. Giving the parents some control over what the doll teaches the child, control they have in other domains such as what media the child has access to, will be crucial. (On the other hand, society will retain an interest in other areas, such as preventing parents from making the doll teach their child bigotry.)

What should the doll report or keep secret? If the doll detects signs of depression, should it tell the parents? What if a child with homophobic parents tells the doll she is gay? What if the doll detects signs of abuse? These are all things modern AIs could be trained to glean from conversation. Reporting depression might help the child, but reporting that the child is gay could hurt her, depending on how such news is received. If the doll reports on abuse to the government, then the doll is actively surveilling its owners; if it does not, then the doll-makers have created a system that can detect abuse, but ignores it. Furthermore, different cultures have different standards of abuse: in many modern cultures striking a child is always considered abusive; in others, corporal punishment is a widely used and socially acceptable form of discipline. Corporal punishment may well fall into the category that we collectively decide is never acceptable. If so, people in societies that routinely use corporal punishment are more likely to be reported if they own the doll: they will be punished for having purchased it. And it is easy to take this benevolent principle to an absurd Orwellian extreme, with the doll being made to encourage a child to report her dissident parents. The tension between encoding the ethics necessary to maintain society and enabling individuals and organizations to teach their AIs their own ethical systems is as much about restraining society to protect individuals as it is about restraining individuals to protect society.

In some domains, current technologies already rely on default strategies that may have moral consequences. Non-prosocial lying is such a case: administrative assistants, for example, may lie about their bosses' availability; should a digital assistant be able to tell such lies? If so, to what degree can the lies be taken? Though not in widespread use, digital secretaries are already being used in the real world (e.g. (Pejsa et al., 2014)), and need an answer to this question. To take another example, should Siri try to stop you from drunk-dialing your ex (or at least argue with you about it)? When should a digital assistant make decisions for you, or try to impact your decision-making?

Towards Ethical AIs

It should now be clear that there is no single morality, either across or within cultures, that could be built into an automated system, and that humans should have some say in their machines' ethics. However, we cannot simply trust humans to make their machines always do the right thing. People convicted of driving drunk have breathalyzers installed in their cars that lock the engine; people with restraining orders could have software installed on their phones to prevent harassment. While in general you have the right to regulate your own behavior, and your technology should help you do that, in some cases, society (or the state) clearly has an overriding interest that limits your behavior. Humans are all too happy to exploit and harm each other, especially when cloaked in anonymity or acting on someone who is not a member of their in-group. It is not only psychopaths or evil people who behave this way; it is well known that the internet can bring out the worst in people (Suler, 2004), and AI should not help them. Again, the pressing issue is neither determining the set of universal morals nor building a fully personalizable ethical system. Rather, the challenge involves navigating the tension between imposing an external ethical system upon people and protecting the interests of society as a whole. The rest of this essay will deal with how to address this problem.

The first step is to collectively determine the set of ethics which *must* be built in to protect us from psychopathic AIs and the worst of human behavior. These ethics will by definition not cover all situations, but only those which society has a fundamental interest in regulating. Existing laws can be the starting point for this discussion: if we have collectively decided humans should not be allowed to do something, computers should not either. Of course, laws trade off against each other, and there are situations wherein otherwise forbidden things are permitted (e.g. violence in self-defense). This process will bring together scholars of the humanities, social sciences, and law to determine this core set of ethical principles, and AI researchers to explain how the realities of implementation will affect the theory of what is being implemented (and to implement it). Everyone involved must understand and communicate that the ethics being agreed upon will not only be regulating systems used by others, but by ourselves, given the problem that (Bonnefon et al., 2016) identify. The ethics defined for autonomous systems may well be different from those governing humans (Hristova & Grinberg, 2016); the point is that we must be clear about what it is we are building, and why.

An approach like Conditional-Preference Nets (CP-Nets, (Greene, Rossi, Tasioulas, Venable, & Williams, 2016)) may work for this task. CP-Nets define what behaviors are preferred under particular conditions. For example, in a situation where a person is in immediate danger, having a robot help that person escape might be preferable to violently defending them, but attacking the aggressor might be permitted if escape is impossible (unless, for example, the aggressor is a law enforcement officer). If the person is being intimidated without direct threat, however, the robot is not allowed to attack the aggressor until the threat becomes tangible. CP-Nets can encode these contextual differences in preferences.

Once that work is done, we must determine a framework within which other ethical concerns can be identified and personalized to the end-user, including the domains, the means, and the extent to which they can be personalized. Such personalization will take time. It will be important to develop a portable standard that can be trained once, with relatively few exposures per principle, and plugged into a variety of systems: people will not want to have to retrain every new device. We will briefly discuss some possible approaches to implement this personalization (see (Malle, Scheutz, & Austerweil, 2017) for a discussion of an autonomous moral agent's desired competencies and qualities).

The most tempting approaches, given modern sensibilities, will be Deep Learning or Reinforcement Learning. Deep Learning learns complex patterns in large feature-rich datasets, and Reinforcement Learning uses reward and punishment to learn to navigate complex state-spaces. These approaches may well work for building in universal moral norms, but not for personalization: both require too much training data for individuals to provide. Some moral situations may only arise once, and it is important to either get them

right, or to correct the wrong behavior, the first time. Furthermore, current implementations of these approaches cannot explain *why* they got an answer. When the system makes a decision the human disagrees with (as it inevitably will), it will need to provide an explanation that the human finds satisfactory, or the human will quickly stop trusting the system.

Another possible approach is to use collaborative filtering, which predicts a user's behavior based on their similarities and differences with other users. However, people might not like being grouped with those with whom they have moral disagreements, even if they agree in some areas. A libertarian and a liberal socialist might agree that the government has no business regulating adults' personal sexual relationships or drug use, but disagree on publicly funded healthcare. Predicting on the basis of the first several shared features that they will share the latter will steer the system wrong. Collaborative filtering also has similar problems to Deep and Reinforcement Learning concerning volume of training data and quality of explanations.

Rule-based systems can incorporate contextual information and readily generate explanations. However, for our purposes rules would have to be learned from a small number of noisy exposures. Rules will only fire if conditions precisely match their triggers, and the system must know how rules trade off with each other (as morals do). In a moral domain with messy real-world data and an enormous number of inputs, these can be significant challenges.

Graphical models such as Bayes-Nets, which encode probabilistic conditional dependencies, may be effective at taking context into account and trading off values. However, Bayesian learning is computationally expensive, especially with high-dimensional input like real-world data, and can require large amounts of input data or a carefully crafted prior.

Finally, Case-Based Reasoning (CBR) has a long history in legal justification and reasoning, and has been investigated as a moral-reasoning technique (e.g. (Blass & Forbus, 2015)). CBR can work from single examples and use whatever information is provided in the case: reasoning is as rich as the input provided and the adaptation technique. Explanations are generated through mapping and adaptation. Experiences and instruction alike can be stored as cases with which to reason, and reasoning can be done on the basis of a partial match. The challenge with a CBR approach is that the system needs to have a relevant case to apply, and needs to know how to adapt it to the current situation.

As to what should be personalizable, a starting place may be decisions that primarily impact the proprietor of the system and their immediate social circle, such as decisions that involve disbursing the proprietor's resources (financial or otherwise). In cases such as hospital management systems and law enforcement support, the proprietors may be groups of people. Of course, anything that conflicts with the established overriding social ethical concerns must be exempted.

Embedding ethical principles into autonomous systems will be a time-consuming and error-prone process. The classic book *I, Robot* illustrates the challenges involved even in simple rules such as "Do what I say except when it harms others". Even assuming we can build the core ethical code into the system before it goes to the end user, our morals are complex, and we should not assume the system has learned our preferences before we've verified that it has. These systems should therefore be designed as apprentices to learn over a long period of time. While they are learning, they should have a less-trusted "trainee" mode, with default behaviors that are explicitly known to be mutable. During the apprenticeship, the system will not be allowed to take most actions without checking with a human first. Humans will understand that the period of time during which their machine nags them will be finite, and that errors are likely. We do not assume that a small child that has made a moral mistake is a psychopath, we correct her; similarly, we should understand that, early in the process, the computer is expected to make mistakes. We need to have a sense of the level at which a system can reason about morality, and trust it to make decisions to the same extent we would trust humans reasoning at the same level. Even after the training period is over, the system should have a confidence threshold below which it will consult a human. And explanation is crucial: if

a system makes a decision with which a user disagrees, but can provide a reasonable explanation, grounded in norms, as to why that decision was made, the user might still accept the system as being ethically competent (and might accept the decision). If the system cannot explain its decision, however, the human will rapidly lose trust in it.

This issue is pressing for several reasons. First, autonomous systems are already making moral decisions, as we have seen. Automated managers are in hospitals; automated assistants are in offices; self-driving cars are on the road; Hello Barbie is in homes. Our phones already help us violate social norms (by letting us drunk-dial); they should also be helping us uphold them. But the bigger reason this issue is of current importance is that this work needs to be done *before* these systems become truly ubiquitous. We need to collectively determine what it is we all agree on and how those common values will trade off with each other before systems are built that simply do it for us. Companies need to know what it is they must (or may) build into their systems. And it will be useful to have a common set of standards that an end-user can train once, then carry from system to system. Fundamentally we are talking about a set of laws and industry standards, and developing those takes time. That must be done now, in advance of pervasive deployment of these systems, rather than attempt to regulate them after they are in widespread use.

Conclusion: An Urgent Frontier

Let us finish with a question barely addressed here but that requires an answer to achieve the above goals: when should systems actively prevent their users from doing things that are illegal, or just wrong? This issue was touched upon in the Hello Barbie example, which points to the beginning of an answer. Certainly in some cases these systems should intervene to prevent harm: automatic braking systems, for example, can already prevent humans from running over people. But in most cases, personal devices should not act as law enforcement. Laws overlap and interact, and enforcing them would require autonomous systems to be legal scholars. People are also unlikely, for example, to buy a car that writes itself parking tickets. If these systems do not enforce laws and regulate illegal behavior, however, they will have to navigate the grey area between *allowing* something to happen and *enabling* it. Is a robot that stands still and allows its user to climb on it in order to crawl through a window participating in a break-in? Is a robotic wheelchair that takes its driver to a drug-dealer helping buy drugs?

Balancing the needs of the group against the freedom of the individual has been long been one of humanity's central projects. With advances in Artificial Intelligence, this old problem moves into new territory. Now is the time to begin navigating the tension between protecting society's interest and empowering people to have systems that reflect their personal convictions. When distinct ethical systems are equally compatible with a safe and well-functioning society, imposing one of them on someone who adheres to another goes against freedoms at the center of pluralistic societies. Whenever possible, we should leave these options open for the users of autonomous systems, while being careful not to give people the power to exploit and oppress others. The line is wide and blurry, and we will need to determine the answers to these questions soon. To simply do nothing is to force this burden upon the programmer, but this is rightfully society's burden to bear.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May). *Machine bias.* Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Blass, J. A., & Forbus, K. D. (2015). Moral decision-making by analogy: Generalizations versus exemplars. In *29th aaai conference on artificial intelligence* (pp. 501–507).

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56–62.

Garvie, C. (2016). *The perpetual line-up: Unregulated police face recognition in america*. Georgetown Law, Center on Privacy & Technology.

Greene, J., Rossi, F., Tasioulas, J., Venable, K. B., & Williams, B. C. (2016). Embedding ethical principles in collective decision support systems. In *Aaai* (pp. 4147–4151).

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, *108*(4), 814.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of personality and social psychology*, *65*(4), 613.

Hristova, E., & Grinberg, M. (2016). Should moral decisions be different for human and artificial cognitive agents? In *38th conf. of the cognitive science society.*

Lee, K., Cameron, C. A., Xu, F., Fu, G., & Board, J. (1997). Chinese and canadian children's evaluations of lying and truth telling: Similarities and dfferences in the context of pro-and antisocial behaviors. *Child development*, *68*(5), 924–934.

Malle, B. F., Scheutz, M., & Austerweil, J. L. (2017). Networks of social and moral norms in human and robot agents. In *A world with robots* (pp. 3–17). Springer.

Pejsa, T., Bohus, D., Cohen, M. F., Saw, C. W., Mahoney, J., & Horvitz, E. (2014). Natural communication about uncertainties in situated interaction. In *Proceedings of the 16th international conference on multimodal interaction* (pp. 283–290).

Sachdeva, S., Singh, P., & Medin, D. (2011). Culture and the quest for universal principles in moral reasoning. *International Journal of Psychology*, *46*(3), 161–176.

Scheutz, M., Malle, B., & Briggs, G. (2015). Towards morally sensitive action selection for autonomous social robots. In *Robot and human interactive communication (ro-man), 2015 24th ieee international symposium on* (pp. 492–497).

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior*, *7*(3), 321–326.

Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, *4*(6), 479.

Joseph Blass is a JD-PhD Candidate studying AI, Cognitive Modeling, and Law at Northwestern University. He is particularly interested in how humans can teach AIs morals and ethics, and how those systems can in turn properly make and justify their moral and ethical decisions.

Sexbots:The Ethical Ramifications of Social Robotics' Dark Side

Christian Wagner (University of Southern California; csondergardwagner@gmail.com)
DOI: 10.1145/3175502.3175513

Introduction

Recently I sat down with an older friend. During our conversation, it came up that I was researching the sex robots ("sexbots") industry. His initial reaction was to laugh in disbelief. "Are you kidding me? You mean like the robots in sci-fi?" My friend's disbelief makes sense; the idea of an actual industry for sex robots seems like a Hollywood movie in the tone of *Ex Machina*. However, the "sexbot industry" is real, and it's growing in popularity. According to a recent study from Tufts University, people find it "relatively permissible" to have sex with a robot. In fact, over 50 percent of people surveyed said they were open to using one (Scheutz, 2016).

While the idea of sex robots has been seen repeatedly in sci-fi, the reality is far removed from what is presented. Current sexbots are not sentient beings ready to please, but rather animatronic sex dolls designed to entice human counterparts through mimicking human appearance, speech, and movement. If interested, one can obtain the latest of these social robots from websites like www.realdolls.com or www.truecompanion.com. I refer to sexbots as social robots, because they are a specific instance of this wider classification of robots.

Social Robots stem from an area of robotics research referred to as Human Robot Interaction (HRI) and are designed specifically to interact with humans on an emotional level. This can take on many forms, from Sony's robotic dog Aibo to home assistant robots like Jibo, a personal AI with a camera and an expressive face (Bartneck et al., 2009; Ackerman, 2017). Despite the vast difference in these robots applications, both share an essential aspect of social robotics: emotional manipulation. Social robotic engineers specifically design their machines to manipulate human emotion in order to cultivate human attachment to a robot, or in other words the robot cultivates a unidirectional emotional connection (Sullins, 2012; Fong et al., 2003; Scheutz, 2009). This emotional manipulation creates a unique ethical dilemma engineers must face: how much emotional manipulation is acceptable (Sullins, 2012)?

Sexbots can be considered social robots, as their creators utilize the same emotional manipulation techniques Jibo and Aibo employ in creating intimate emotional bonds with human users. As such, the sexbot industry must face not only the ethical questions raised by being part of the sex industry but also those faced by the whole of the social robotics. If ethicality of social robots, in general, can be considered ambiguous and in need of further discussion, then the need for examining the ethicality of sexbots is pressing. For Sexbots pose a grave threat to the public in various ways. Not the least of which is the ethical danger all social robots pose: power to manipulate human emotion.

Social Robots Emotional Influence

If the idea of having strong emotional bonds with a machine seems far-fetched, consider the strong attachment adults exhibit towards personal pets or service animals. For example, on a cold January morning in Cleveland, hundreds gathered to pay their respects to a fallen police dog, Jethro, who was shot apprehending a criminal (Loreno and Gallek, 2016). Jethro served the public valiantly, creating strong emotional relationships with his co-workers, despite the fact that he wasn't a human. Humanity is gifted with the ability to form relationships; it is a gift robotic researchers, and the United States military, have recognized.

Soldiers stare in solemn silence at the remains of their fallen comrade. They recall all the times "Scooby" saved their lives from enemy IEDs. They recall the time they introduced Scooby to their families stateside and the time he first joined the team. It didn't matter that Scooby was just a robot, he was family. While this story is fictitious, the details are all true. When the military assigned robots to bomb

disposal teams, it likely did not anticipate the extent of soldiers' relationships with the bots. Yet studies have found that soldiers create intense emotional bonds with their IED disposal robots. The soldiers name them, introduce them to families, and even hold funerals after they are destroyed (Scheutz, 2009). This emotional attachment is not limited to combat. Consider, for example, how effective IRobot's Roomba is at tugging domestic heartstrings.

Surveys of Roomba owners found that approximately "two-thirds had named their devices and half had assigned them arbitrary genders" (Dillow, 2010, para. 2). Roomba owners have gone as far as rearranging their houses to accommodate their robotic companions (Scheutz, 2009). Some even clean up the house themselves to give the robot a well-deserved break (Dillow, 2010). The Roomba is just a vacuum cleaner, employing extremely few of the vast number of features that HRI research has determined naturally lead humans to form deep emotional bonds.

The HRI community has over the years accumulated large amounts of research discerning exactly how to design and program robots to hack our psychological makeup. Their goal is to create fluid human robot interaction by exploiting humans' emotional tendencies to form relationships with entities of natural or perceived intelligence (Sullins, 2012; Bartneck et al., 2009; Fong et al., 2003). While "exploiting emotions" may sound sinister, HRI research is primarily focused on the betterment of society. Research goals include: creating robots to improve the social skills of children with autism, provide health-care and companionship to the elderly and infirm, and improve human interactions with robots as a whole (Scassellati, 2007; Broekens et al., 2009). So, while many of these tasks require a level of persuasion and emotional influence to achieve their goal, they seem ethical in light of their altruistic goals.

Social robots influence our emotions through a number of different features. Researchers give robots faces, eyes, and animal-like appearances, then program the robots to have personalities and respond to human stimuli like motion and verbal cues. By employing these features in robots like Jibo and Aibo, roboticists are finding they can consistently enhance humans' emotional attachments to their machines (Fong et al., 2003; Bartneck et al., 2009). And as the research continues, social robots are becoming more emotionally persuasive at a staggering rate (Scheutz, 2009; Ackerman, 2017). This reality should cause us to join roboticists and tech-savvy individuals in thinking about the ethical dangers social robots present. Although they may look cute and innocuous now, social robots of the future have the potential to manipulate the emotions of humanity on an unprecedented scale with less altruistic goals than their predecessors (Bartneck et al., 2009).

By designing robots with the distinct functionality to manipulate our emotions, engineers are creating a mechanism by which entities can perform emotional extortion. For instance, a company can use an individual's emotional bond to subtly influence buying patterns. Less subtly, engineers can program the robot to threaten to end the relationship unless its owner buys them a new accessory (Scheutz, 2009). Either way, the issue is the same: is it ethical for any social robot to "play on deep-seated human psychological weaknesses put there by evolutionary pressure" (Sullins, 2012, p. 408)?

The social robot community as a whole has recognized this ethical pitfall and has proactively started ethical discussions, going as far asking for legal entities to protect the public from emotional extortion from their robotic creations (Scheutz, 2009). Such legal actions will aid in keeping social robots' intentions altruistic (as in helping children with autism), or neutral (like the Roomba). Because of these discussions and legal actions, the full judgment on the ethicality of social robots as a whole remains to be seen. However, the emerging sexbot industry continues to remain apart from the general HRI and social robot community, meaning they are not engaging in these needed discussions (Scheutz, 2016). A troubling fact.

Sexbots, Potential For Harm

Sexbots pose their own unique dangers, by putting those who form strong emotional bonds with them at risk of emotional extortion at the hands of a largely unethical industry looking only to make money. The compa-

nies who build the robots have the means to program it to use any tactic they want, manipulating or outright extorting users into spending more money on their industry. And it will work. It will work because "Sex Sells" and people are willing to pay to continue/advance the experience. Marketing companies have known "for over 100 years" that "[w]hen ads are more sexually provocative, men in particular are irresistibly drawn to them" (Suggett, 2017, para. 8). For example, after the movie 50 Shades of Grey came out, a "British sex toy retailer Lovehoney saw a 30 percent increase in sales" while another saw a "25 percent increase" (Hanlon, 2017; Kharpal, 2015, para. 2; para. 7). With claims and statistics like these, it appears sexual imagery already has enough power to sway the public's decisions. Now imagine how coerced into spending money a person would be if their sexbot, that they have an intimate sexual emotional bond with, told them to buy something, as opposed to a TV add telling them.

In addition to the potency of their products, sexbot companies could employ a marketing scheme in which users pay to enhance or continue the experience. This kind of marketing is often seen in the video game industry with mobile games employing "pay to play" and mainstream games employing Downloadable Content (DLC) where users pay to gain access to extra game content (Sinicki). The "pay to play" marketing scheme is analogous to old school arcade games, where the player pays a few cents to keep the fun going. The mobile game Candy Crush is infamous for sucking money out of their players, making "$800,000 daily" from millions of users paying to continue advancing in the puzzle game (Smith, 2014, para. 2). As for DLC, Electronic Arts (EA) made an estimated "$1.3 billion" off DLC alone in 2015 (Thier, 2016, para. 2). If people are willing to pay so much to advance in a mobile game and for video game DLC, how much more will people spend on DLC to increase their sexbot's vocabulary or expand its sexual repertoire. This puts immense power into the hands of the sexbot industry and puts the public at risk of excessive emotional manipulation. All social robots could employ this kind of marketing; however, sexbots pose an additional threat with the added element of sex and the negative psychological effects it can cause.

Because there is little empirical data on sexbots' psychological effects on users, I will compare them to another inanimate source of sexual gratification: pornography. The effects of pornography have been extensively studied, and there is a continually heated debate raging around whether or not pornography is actually harmful to those who view it (Hald and Malamuth, 2007). Some researchers find pornography's negative effects to be relatively small and outweighed by positive effects. They find that pornography can serve as a source of sexual information and lead to increased sexual experimentation, correlating with "improved sexual communication [and] enhanced couple intimacy" (Fisher et al., 2017, p. 1). Other researchers say that pornography users experience negative effects to a significant degree: arguing that pornography use can lead to the objectification of others, increased sexual violence, reduced empathy for sexual victims, and damaged relationships (Flood, 2009; Ybarra et al., 2010). Ultimately the debate is a matter of degree; even pornography's advocates admit that pornography use overall manifests at least some negative effects.

That pornography consistently manifests negative externalities is a strong reason for concern when it comes to the topic of sexbots. If pornography, or the viewing of sexually explicit material, causes adverse effects; it is reasonable to assume that the use of sexbots, which physically manifest explicit material and are specifically designed to be emotional influences, would cause proportionally greater effects. Put simply, the greater a user's engagement with sexual material, the greater the influence it holds, a concept that has been shown true for pornography alone (Flood, 2009).

The potential danger of increased sexual engagement is made clearer when you consider that sexbots could be used to gratify the most base and vile of sexual urges. Researchers have found that repeated exposure to sexually violent or pedophilic pornographic material can erode empathy and increase aggression towards real people (Flood, 2009; Ybarra et al., 2010). One study found that intentional viewing of violent pornographic material

can lead to a 6-fold increase in the likelihood of the viewer self-reporting aggressive sexual behavior (Ybarra et al., 2010). Even viewing of moderately sexual material has been shown to correlate with an individual's acceptance to "force a girl to have sex" (Flood, 2009, p. 393). Of course, these kinds of correlations are cyclical by nature, with those viewing the material to be among those most likely to practice it in real life. However, even if a person is predisposed to sexual aggression, engaging with the corresponding pornographic material will further exacerbate and encourage those predispositions. Thus it is very troubling to read a study performed at Tufts University finding that people rank the idea of using a sexbot to engage in "rough sex or sadistic behavior" as a 5.23 out of 7 for acceptable behavior (Scheutz, 2016, p. 354). (On a scale where a 1 is unacceptable and a 7 is completely acceptable).

As of yet, no studies have been performed to correlate these kind of tendencies with sexbots; however, "[l]eading psychologists and social scientists studying this technology argue that sex robots will most likely contribute to psychological disorders rather than mitigating them" (Sullins, 2012, p. 402). This is reasonable. Sexbots are social robots engineered to provide emotional stimulation catered to a user. They are programmed to fully entrance users, exhibiting any physical characteristic and engaging in any behavior to provide the user with an analog to a real-life counterpart. The sexbot industry is already catering to troubling behaviors. One individual designed a sexbot to resemble Scarlett Johansson (O'Neil, 2016). Even more disturbing, a Japanese company, Aibo, allows users so-inclined to order sexbots "resembling five year old girls" (Richardson, 2016b, p. 48).

Providing a legal outlet to practice illegal sexual acts could seem like a potential benefit of sexbots. Some researchers argue that they can serve a valuable role by providing an outlet for those wishing to perform illicit sexual acts, like sex with children (Mackenzie, 2014). However, this line of reasoning has not proven true in the case of sexually aggressive pornography (Flood, 2009). When the alternative is not passive (as with pornography) but takes the form of social robots designed to enable and emotionally influence the user, it seems unlikely that sexbots will suppress people's inclinations. On the contrary, they seem likely to enhance them. Thus the potential benefit of sexbots must be weighed against their potential to erode an individual's empathy towards the real act.

The Potential Benefits of Sexbots

Treating the symptom and not the illness is a concept that can be applied to a variety of things from medicine to the Caped Crusader. The idea is simple. By running around and beating up thugs, Batman might only be treating the symptoms of the much greater illness of systemic crime. Although it is admirable to treat the symptoms when the illness is itself too large to tackle, it is only beneficial if the treatment doesn't enlarge the disease after the initial symptoms are temporarily relieved. Batman throwing another thug in jail doesn't help if it allows the Joker to kidnap the mayor. Most of the arguments for sexbots, fall into this category. Those in favor of sexbots, such as David Levy, the influential author of *Love and Sex with Robots*, argue that sexbots could provide a sexual outlet for those with disabilities and a means of safe sexual education while reducing the demand for sex slaves and prostitutes (Richardson, 2016a; Mackenzie, 2014).

Proponents of sexbots argue that these social robots can be used as substitutes for sex workers, driving down the demand for sex slaves and prostitutes (Mackenzie, 2014). The sex slave industry is not just some fiction like the movie *Taken*. According to CNN, the sex industry enslaves an estimate of "10-30 million" people across the world, from Los Angeles to Bangkok (Tanneeru, 2011, para. 2). If sexbots could fight against this industry, then they should be considered. However, they most likely won't. One reason for this is the practical matter of cost. Sexbots are expensive, so much so that sex slaves are cheap in comparison (cnn, 2013). For instance, a sexbot from www.truecompanion.com costs around ten thousand dollars, whereas a young girl can be bought from Islamic State fighters in Turkey and Jordan for "$124", according to a UN Official (Yoon, 2015, para. 5). Combine this with the fact that the countries with the largest populations of sex slaves (Bangladesh

and Myanmar for example) are relatively impoverished, and it seems unlikely sexbots will have a large impact on the sex slave industry (Karuga, 2016). Even though sexbots are expected to reduce in cost with time, given the advances in robotics the cost is unlikely to decrease a hundredfold (Mackenzie, 2014).

As for the prostitution market, again sexbots seems unlikely to have much of a positive influence. Many prostitutes choose their line of work. According to one survey, "68 percent (of prostitutes) consider their line of work as part of their sexuality" and 85 percent say money is a driving factor for them (Karkov, 2012, para. 1). Applying basic economics, if sex robots reduce the demand for prostitution, it will only drive the cost of prostitution down, thereby harming those who rely on the income to buy "food and day care for their children" and making prostitution more accessible (Karkov, 2012, para. 13).

In addition to these economic arguments for why sexbots won't fight the disease of illegal sex industries, it is important to remember that sexbots are social robots with all the inherent emotional powers they employ. Even if sexbots were to overcome the economic obstacles, swaying individuals to utilize them instead of engaging in prostitution or buying sex slaves, they still present the psychological dangers of being a powerful emotional manipulator involved in an emotional activity. As I mentioned before, sexbots' strong emotional influence can potentially erode the empathy of the user to greater a degree than sexually aggressive pornography. Thus sexbots may encourage sexual malfeasance, thereby growing the sex slave industry. Further, it is unreasonable to assume that sexbots will be used as a substitute for, and not in conjunction with, human companions. The Tufts University study on sexbots found that people ranked "mixed human-robot group sex" a 5.16 out of 7 for appropriate behavior (Scheutz, 2016, p. 354).

The idea to use sexbots as a means for sexual education and to help those with disabilities have a safe positive experience has real potential for good (Sullins, 2012). However, Sexbots' power of emotional influence, inherent in their status as social robots, is an argument against the idea of using sexbots for these purposes. Exposing emotionally susceptible teens to a social robot in an emotional activity is too large of a risk, especially when the alternative is having mature adults talk to teens about sexuality. There are simply better tools than sexbots for the purposes of education.

As for those with disabilities that prevent them from having real and positive sexual relationships, sexbots have the potential for positive use if used properly. However, allowing widespread use of sexbots for this singular good would be like using morphine for a headache; a drastic over-medication, likely to cause more problems than it would solve. If the use of sexbots could be controlled and reserved for those with disabilities, similar to using social robots for children with autism, then they could prove altruistic. But this sort of controlled use has yet to be indicated by the companies making these sex social robots.

Conclusion

In a classic example of the whole being greater than the sum of its parts, sex robots combine social robotics' powerful emotional influencing and the negative effects of the sex industry into a dangerous product that threatens public well-being, thereby making them unethical. The social robotics community as a whole faces the ethical question of how much emotional manipulation is acceptable and in what context. Should the full power of social robotics be reserved for causes like helping children with autism, or is it acceptable to be in everyday machines like Jibo? This discussion is necessary and should continue into the future; however not having a clear answer on the ethicality of social robotics does not prevent answering the ethical questions raised by sex robots.

With the powers given them by social robotics research, sexbots threaten to cause widespread harm to all those who use them. One survey found that most people view using a sexbot on par with masturbating and are open to using one, implying that in the future sexbots could be used by a large percentage of the public (Scheutz, 2016). With this potential to gain popularity it is important to take into account the harm and relatively little benefit of sexbots. The harm includes the expectation of amplifying sexual practices of violence

and pedophilia by enabling repeatable emotion influencing exposure, which erodes empathy over time and potentially leads to unspeakable sexual acts. Sexbots also put into the hands of an unethical industry the means by which to emotionally extort and manipulate the public. This potential for harm is not mitigated by the supposed benefits sexbots promise. For it is highly unlikely sexbots will hurt the sex slave industry and it could even make a prostitute's life harder, simply due to practical economic principles. Sexbots are more likely to actually cause these industries to grow by means of the harmful effects mentioned before.

In considering whether or not sex robots should be created, examine IEEE's code of ethics for engineers which begins, "To accept responsibility in making decisions consistent with the safety, health, and welfare of the public" (IEEE, para. 2). In light of this foremost ethical principle, to hold paramount the public safety, it is the responsibility of companies and individuals to carefully consider the potential harm the sexbot industry poses to public welfare if it continues on its current course. We currently stand upon the edge of what could be the biggest change in the sex industry since the internet, as such the need for ethical discussion is pressing. All those with the technical expertise to understand the full scope of the situation should strive to educate the general public about the potential pitfalls of sex robots.

References

(2013). The number.

Ackerman, E. (2017). Ces 2017: Why every social robot at ces looks alike. *IEEE Spectrum*.

Bartneck, C., Kulic, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1:71–81.

Broekens, J., Heerink, M., and Rosendal, H. (2009). Assistive social robots in elderly care: a review. *Gerontechnology*, 8(2):94–103.

Dillow, C. (2010). Emotional attachment to roombas suggests humans can love their 'bots. *Popular Science*.

Fisher, W., Kohut, T., Graham, S. M., and Campbell, L. (2017). 398 effects of pornography use on the couple relationship: Results of bottom-up, participant-informed, qualitative and quantitative research. *The Journal of Sexual Medicine*, 14(1).

Flood, M. (2009). The harms of pornography exposure among children and young people. *Child Abuse Review*, 18(6).

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42:143–166.

Hald, G. M. and Malamuth, N. M. (2007). Self-perceived effects of pornography consumption. *Archives of Sexual Behavior*, 37(4):614625.

Hanlon, J. (2017). Fifty shades darker release sees spike in sex toy sales. *Inquistir*.

IEEE. Ieee ieee code of ethics.

Karkov, R. (2012). What drives a prostitute. *Science Nordic*.

Karuga, J. (2016). Countries with the most modern slaves today.

Kharpal, A. (2015). Kinky! the saucy business sparked by 50 shades. *CNBC*.

Loreno, D. and Gallek, P. (2016). Public memorial service held for fallen canton k-9 officer jethro. *FOX 8 Cleveland*.

Mackenzie, R. (2014). Sexbots. *Proceedings of the 2014 Workshops on Advances in Computer Entertainment Conference - ACE 14 Workshops*.

O'Neil, K. (2016). Man builds 'scarlett johansson' robot from scratch to 'fulfil childhood dream' - and it's scarily lifelike. *Mirror*.

Richardson, K. (2016a). The asymmetrical relationship: parallels between prostitution and the development of sex robots. *ACM SIGCAS Computers and Society*, 45(3):290–293.

Richardson, K. (2016b). Slavery, the prostituted, and the rights of machines. *IEEE Technology And Science Magazine*.

Scassellati, B. (2007). How social robots will help us to diagnose, treat, and understand autism. *Springer Tracts in Advanced Robotics*, 28:552–563.

Scheutz, M. (2009). The inherent dangers of unidirectional emotional bonds between humans and social robots. *IEEE Intl Conf. Robotics and Automation*.

Scheutz, M. (2016). Are we ready for sex robots? In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 125 – 132.

Sinicki, A. How candy crush is controlling your brain the psychology behind addictive computer games.

Smith, D. (2014). This is what candy crush saga does to your brain. *The Guardian*.

Suggett, P. (2017). Can sexual imagery really drive sales?

Sullins, J. (2012). Robots, love, and sex: The ethics of building a love machine. *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, 3(4):396–409.

Tanneeru, M. (2011). The challenges of counting a 'hidden population'. *CNN*.

Thier, D. (2016). Ea is making a giant amount of money off microtransactions. *Forbes*.

Ybarra, M. L., Mitchell, K. J., Hamburger, M., Diener-West, M., and Leaf, P. J. (2010). X-rated material and perpetration of sexually aggressive behavior among children and adolescents: is there a link? *Aggressive Behavior*, 37(1):118.

Yoon, S. (2015). Islamic state circulates sex slave price list. *Bloomberg Buisness*.

Christian Wagner may be contacted at csondergardwagner@gmail.com. He is a recent graduate from the University of Southern California and Gordon College. Christian is currently working as an engineer for Northrop Grumman Corp. He intends to pursue a graduate degree in robotics.

Automation Moderation: Finding Symbiosis with Anti-Human Technology

Jack Bandy (University of Kentucky; jgba225@g.uky.edu)
DOI: 10.1145/3175502.3175514

Abstract

I examine Artificial Intelligence and automation technology from the perspective of Aristotelian virtue ethics, focusing on the detrimental impacts automation technology can have on human creativity, learning, and attention. As a potential solution, I put forward "automation moderation," which is a human-centered approach to designing AI, and showcase some examples of it.

Introduction

> *"The time has come for computer professionals. We now have the power to alter the state of the world fundamentally and in a way conducive to life."*
> –Joseph Weizenbaum, 1987

Framing automation as an economic dilemma rather than an ethical dilemma is a category mistake that must be soon admitted and corrected. Especially in excess, automation technology exhibits anti-human tendencies, such as degrading effects on our creative capacity, our learning potential, and our ability to focus. To examine automation as an ethical issue, it will first be helpful to look at certain examples of automation technology and to consider the dangers they pose. With these dangers in mind, we can then appreciate examples of automation technology that benefit our minds and bodies, as well as reflect on the distinctions between the harmful examples and the beneficial examples.

In pointing out that job automation technology has not received attention as an ethical dilemma – unlike the self-driving trolley problem, which seems to get attention from every armchair philosopher who has an internet blog – we should not overlook that the technology has received plenty of attention as an economic dilemma. Many have observed the looming threats of unemployment, massive workforce transitions, and other potential consequences of automation technology, mostly through the lens of abstract statistical projections. Gigabytes of unemployment metrics, job reports, GDP growth, and the like have served as the basis of evaluation thus far (Holdren, 2016). But a proper evaluation of automation technology should begin with humanity, not statistics. In other words, we should give attention to both the intrinsic value and consequential value of the technology, but we must start with the former.

Work is a fundamental part of human life, and so to take away someone's work is to take away part of their humanity. Countless civilizations have defended the necessity, privilege, and meaning of labor – though it is worth pointing out that, by historical accounts, the resistance to work in modern western society is quite bizarre. It is possible that the communities of artificial intelligence researchers and automation experts know the importance of work, and merely want to "save us" from expending effort on the unimportant things – "the things nobody wants to do" – but a few examples will show that is not the case. In fact, we will see that unmoderated automation technology has already corroded human creativity, learning and attention.

Cold creativity

Although manual labor jobs may come to mind first when discussing job automation, a great deal of automation technology actually aims for human creative processes. One example is architectural design. Given the rich history of cultural expression that one can learn from studying architecture, it is surprising the extent to which computer-aided design (CAD) programs have sought to expunge humans from the design process. This was not always the case. The first CAD systems encouraged creativity, giving the designer a new set of expressive tools, all the while improving architectural

precision. But today, mostly as a result of CAD software that accepts a handful of parameters as input and provides a full building design as output, architecture is becoming less and less a creative and expressive human endeavor. Instead, "parametricism" is the defining post-modern architectural style, described by one renowned architect as a style marked by "blobs" that can be produced with a few clicks on the computer (Schumacher, 2010). With an excess of automation, CAD programs have now put the computer, in its cold and mechanical decisions, at the center of a human design process, that is, CAD programs have "saved us" from the creative process. We will see later that these automation efforts in moderation can empower human creativity while still capitalizing on the power and precision offered by the computer.

Lax Learning

Automation excess has also found its way into the aviation industry. Just as some CAD programs remove humans from design decisions, some flight systems remove humans from navigation decisions. Airbus has indicated that it hopes to make planes that are "pilot-proof," with humans sitting in the cockpit essentially to babysit the aircraft. On the surface, this form of automation seems like an innocent effort to cut down on labor. It takes diligent time and practice to learn the rules of flight, many hours to gain certification, and still many more hours at the yoke to truly master the complexities of aeronautics. How could cutting out this effort be such a bad thing? From an alarmist perspective, one could point to autopilot, and specifically, to Air France flight 447, which crashed in the Atlantic ocean and killed each of its 228 passengers in June of 2009. Part of Airbus' automation-centered design is that mechanical yokes are replaced with what are essentially digital joysticks. In traditional planes, both pilots hold their own yoke, but they are mechanically linked such that both yokes move identically. On flight 447, this was not the case, and black-box data has revealed that one pilot was pulling back on his joystick, a critical (and in this case fatal) error which the other pilot would have noticed had they been steering with mechanically linked yokes instead the joysticks that supposedly "save them" physical energy. Automation-centered design also encourages excessive use of the autopilot feature, which prevents pilots from practicing and maintaining the many skills required to fly an aircraft. In situations like 447, practicing could be the difference between life and death.

But besides this warning against automation-driven flight, there is reason to embrace human-centered flight. It is arguably the very reason the Wright brothers were successful: an intrinsic appreciation for the wonderful experiences that one finds as a result of learning efforts. Once the hours of learning are applied, however arduous that learning may be, applying this knowledge to partake in human flight is nothing short of bliss. Thus, to shortchange the work of learning is to shortchange a deep, hard-earned, blissful satisfaction.

Dull Decisions

Finally, in addition to automation technologies that degrade the mind's creativity and learning power, automation excess has also threatened our attention and focus. A clear is example is the "app suggestion" feature on both iOS and Android. This feature prompts a user to open an app based on the user's habits combined with current context details such as location, time of day, and connected devices. Just as CAD originally enabled creativity and the plane originally offered bliss from learning, the phone originally enabled focused attention. In the days of the payphone booth, one had to make the active decision to spend their attention on a phone call to a specific person for a specific amount of time. Now, with automation features such as app suggestions, it is the phone that makes the active decision, using probabilistic models and statistical learning to tell the user what he or she should do with precious time and attention. One may argue that app suggestions save the user time, but few of us would consciously choose the few seconds it takes to make a few extra taps over our very autonomy – our will and our ability to make choices are both essential to our humanity, and must not be automated. In a proverbial sense, we were once able to use the phone, but automation means the phone is able to use us. We should be weary of technologies that promote this reversal, especially if they only "save us" a couple

of seconds.

At this point, we can move beyond the gloominess of unmoderated automation technology, and examine what might be called "automation moderation." In short, creating this type of automation technology simply asks that we discern between "can" and "should" before putting automation technology in the center of human endeavors.

Semi-automatic

Although many CAD programs are drifting towards fully automated design processes, some have recognized the need for human creativity and already offered robust, human-centered solutions. For example, some researchers have designed a CAD program (Gross & Do, 1996) that allows for freehand drawing as an input mode. This promotes a designer?s connectedness to the design, partly because moving a pen with your hand engages more sensory perception and cognition than tapping a keyboard. A blank page is also more conducive to creativity than a keyboard – a keyboard limits possibilities, a blank page leaves possibilities to the designer. All the while, this program uses sophisticated refinement processes to discern the designer's intention and thus capture it with the detail and precision required for modern manufacturing. Whereas removing humans from this process discourages them from creative design, putting humans at the center of the creative process preserves and encourages their capacity for innovation.

Changing times

No discussion of automation would be complete without mentioning the assembly line. For a short case study, the Detroit-based watch manufacturing company Shinola will be useful. Although assembly lines are generally not known for their teaching abilities, Shinola defies that reputation by focusing on "skill at scale," which involves hand-making each individual piece of the watch from start to finish. Workers learn different parts of the artisanal process as they spend more time at the company under an apprentice system, a system that encourages workers to gain a skill set rather than encouraging them to watch a machine remove the need for that skill set. Although not all manufacturing can use humans instead of machines from start to finish, in this case, humans are properly valued for their ability to learn and perform a skill in a way that automated machines cannot. A machine can follow a recipe, but only a human can taste-test.

A phone of the people, for the people, by the people

Lastly, an example of automation moderation that gives you back your attention and fits in your pocket (Siempo, 2017). Marketed as "the phone for humans," Siempo is designed from the ground-up with human psychology as its main consideration. Rather than suggesting apps and behaviors, Siempo's interface is dubbed the "intention field," where users must type in the course of action they wish to take using their phone. Instead of bombarding users with unpredictable alerts that vaporize attention, Siempo sends notifications during windows that the user defines, for example, once every hour. This puts decision-making abilities back into the hands of the user, avoiding the incessant and unpredictable interruptions that characterize the modern smartphone. In other words, this human-centered phone is designed to allow and encourage users to practice focusing their attention. Like creativity and learning, the ability to focus our attention is an intrinsically valuable skill, and one that any human will find worthwhile in and of itself. Siempo recognizes the importance of attention, while still giving users the basic tools and technology that we now expect from our phones.

Conclusion

As this essay has shown, automation can not be reduced to statistics, whether the decimal points represent income dollars or employment percentages. Unfortunately, the popular way to construe automation technology and job loss is merely as a money issue. When construed as a money issue, only money solutions are needed. And so, many technologists (including Elon Musk) have unabashedly suggested universal basic income as a solution to the dilemmas of automation technology. But the issue, which is ethical before

it is economical, is that automation excess causes people to lose creativity, learning capacity, and ability to focus. These are qualitative, human-centered concerns rather than quantitative, mechanically-centered concerns. If we aim to solve the quantitative concerns, both people and statistics will suffer. Only if we aim to solve qualitative concerns do we have a chance, in the long run, to mend both.

Acknowledgments

I am thankful to Dr. Emanuelle Burton, whose class helped me practice thinking and writing about these topics with greater descriptive clarity. I am also grateful to Dr. Judy Goldsmith for showing us the opportunity to enter the SIGAI essay contest. Also, thanks to my mom for her skillful editing as I prepared the final version of this paper.

References

Gross, M., & Do, E. (1996). *Ambiguous intentions: a paper-like interface for creative design.* https://wiki.cc.gatech.edu/designcomp/images/archive/e/e4/20110701215646!Uist96-ambiguous.pdf.

Holdren, J. P. (2016). *Preparing for the future of artificial intelligence.* https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

Schumacher, P. (2010). *Patrik schumacher on parametricism - 'let the style wars begin'.* https://www.architectsjournal.co.uk/the-critics/patrik-schumacher-on-parametricism-let-the-style-wars-begin/5217211.article.

Siempo. (2017). *The phone for humans.* https://www.kickstarter.com/projects/siempo/the-phone-for-humans.

Jack Bandy studies Computer Science in the graduate program at the University of Kentucky. He hopes to finish his master's degree in May 2018 and then find a job or Ph.D. program that helps ensure machine learning works for holistic human flourishing.

AI Conference Reports

Michael Rovatsos (University of Edinburgh; mrovatso@inf.ed.ac.uk)
DOI: 10.1145/3175502.3175515

This section features brief reports from recent events sponsored or run in cooperation with ACM SIGAI.

The 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017)

Sao Paulo, Brazil, May 8-12, 2017
www.aamas2017.org
AAMAS 2017 was a success overall, with a strong technical program and a healthy number of 410 attendees. The first two days of the conference were dedicated to workshops, tutorials and the doctoral consortium. There were 18 workshops (2 two-day workshops, 10 one-day workshops, and 6 half-day workshops) and 9 tutorials (all half-day). The main conference program ran from May 10-12. The main conference featured 5 sessions with 6 parallel tracks in each session, 4 poster and demonstration sessions, 3 keynotes by Ana Bazzan (Federal University of Rio Grande do Sul, Brazil), Julie Shah (MIT), and Jeff Schneider (Uber ATG and CMU), and two award talks by David Parkes (Harvard) and Nisarg Shah (University of Toronto). The general technical quality of the papers was high, and certainly comparable to previous AAMAS conferences. In total, 567 papers were submitted for review. Of these papers, 155 were accepted for presentation as full papers. This resulted in an acceptance rate of 27%. The workshop and tutorial program was well attended with 74% of conference participants attending at least one day of the workshop/tutorial program. The next edition of the conference, AAMAS 2018, will be held is part of the Federated AI Meeting (in Stockholm, Sweden, July 10-15, 2018.

International Joint Conference on Rules and Reasoning (RuleML+RR 2017) *London, UK, July 12-15, 2017*

2017.ruleml-rr.org/
In 2017 the two events RuleML (International Web Rule Symposium) and RR (Reasoning and Rule Systems) came together to form the first RuleML+RR event. The event was co-located with the 13th Reasoning Web Summer School, DecisionCAMP, and the 31st British International Conference on Databases, and was attended by 122 delegates. RuleML+ RR 2017 provided a rich programme of four tutorials, three keynote talks, and technical paper sessions, together with several sub-events, including a Doctoral Consortium, the 11th International Rule Challenge, an Industry Track, and a Poster session. In total the conference included 14 research papers selected from 38 submissions. The Rule Challenge consisted of 5 presentations and demos, in diverse areas, such as traffic management, health insurance, diabetes counselling support, anomaly detection, and translation from (controlled) English to rules. As well as the Doctoral Consortium best paper award, there were awards for the best paper, the Rule Challenge, and the best poster. The conference included a fascinating after dinner speech given by Bob Kowalski (Imperial College London) titled “Logic and AI – The Last 50 Years” which can be viewed online. The next edition of the event, RuleML+RR 2018, will be held in Luxembourg, September 18-21, 2018.

The 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017)

Porto, Portugal, February 21-23, 2017
www.biostec.org/?y=2017
BIOSTEC 2017 received 312 paper submissions from 56 countries, of which 71 papers were accepted as full papers and 128 papers were accepted as short papers (71 with oral presentation and 57 with poster presentation). The resulting “full?paper” acceptance rate of about 23% and the total oral presentation acceptance rate close to 45% demonstrate the organizers’ intention of preserving a high quality forum for the next editions of this conference. The program included four invited talks delivered by internationally

distinguished speakers: Bart M. ter Haar Romeny, (Eindhoven University of Technology, Netherlands), Kristina Höök (Royal Institute of Technology, Sweden), Bethany Bracken (Charles River Analytics, United States), Hugo Plácido da Silva (Institute of Telecommunications and PLUX Wireless Biosignals, Portugal). BIOSTEC 2017 also featured a panel session on "Biomedical's future: Do we need a SWOT analysis?" and a European Project Space session on "Biomedical Systems and Technologies". The next edition of the conference, BIOSTEC 2018, will be held during January 19-21 in Funchal, Madeira, Portugal.

The 30th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE-2017)

Arras, France, June 25–30, 2017
www.cril.univ-artois.fr/ieaaie2017/
IEA/AIE-2017 continued the tradition of previous events in terms of emphasizing research advances on new and innovative intelligent systems methodologies and their applications in solving complex real life problems. In this 30th year, we received 180 papers for the main and special tracks, and accepted 70 papers as full papers and 45 papers as short papers. The conference greatly benefited from invited talks by three world-leading researchers in applied artificial intelligence, Jian J. Zhang (Bournemouth University, UK), Umberto Straccia (CNR-ISTI, Italy), and Leendert van der Torre (University of Luxembourg). In addition to the main track, the conference included special tracks on Agronomy and Artificial Intelligence, Anomaly Detection, Applications of Argumentation, Conditionals and Non-monotonic Reasoning, De Finettis Heritage in Uncertainty and Decision-Making, Graphical Models: From Theory to Applications, Innovative Applications of Textual Analysis Based on AI, and Intelligent Systems in Health Care and mHealth for Health Outcomes. IEA/AIE-2017 also organized nine poster sessions, and two workshops, one on ASP Technologies for Querying Large-Scale Multiple-Source Heterogeneous Web Information, and one on Computer Animation and Artificial Intelligence.

The 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)

Porto, Portugal, February 24-26, 2017
www.icaart.org/?y=2017
ICAART 2017 received 158 paper submissions from 45 countries of which 42 papers were accepted as full papers and 75 papers as short papers (35 for oral presentation and 40 posters). The resulting full paper acceptance rate of about 26% and the total oral paper presentation acceptance rate close to 50% show the intention of preserving a high quality forum for the next editions of this conference. Apart from paper and poster presentations, the ICAART 2017 program included four invited talks delivered by internationally distinguished speakers, Vanessa Evers (University of Twente, Netherlands), João Leite (Universidade Nova de Lisboa, Portugal), Nuno Lau (Universidade de Aveiro, Portugal), and Francesca Rossi (IBM, United States and University of Padova, Italy). ICAART 2017 also featured a panel session on "How Computers View and Measure Ethical Decisions by Intelligent People" and a European Project Space Panel on "Intelligent Systems and Technologies". The next edition of the conference, ICAART 2018, will be held in conjunction with ICPRAM 2018 in Funchal, Madeira, Portugal, from February 16 to 18, 2018.

Michael Rovatsos is the Conference Coordination Officer for ACM SIGAI, and a faculty member of the School of Informatics at the University of Edinburgh, UK. His research in in multiagent systems and human-friendly AI. Contact him at mrovatso@inf.ed.ac.uk.