COMPUTING edge

- Security and Privacy
- Data
- Internet
- Artificial Intelligence

OCTOBER 2020

www.computer.org

IEEE COMPUTER SOCIETY

IEEE

Thank You to Our Sponsors!

CVPR 2020—the premier annual conference on computer vision and pattern recognition—virtually delivered more than 5,000 first-class papers, keynotes, sessions, workshops, and tutorials to an audience of 7,600 from all over the world!

We are so grateful to our wonderful volunteers and our amazing lineup of sponsors for their continuing support.

Champion

Alibaba Group 阿里巴巴集团 · amazon | science · appen · MEGVII 旷视 · Microsoft · NAVER | LINE | LABS · Qualcomm · SuperAnnotate

Supporter

ALBEVERIE · algolux · Baidu 百度 · EXXACT · FACEBOOK · FUTUREWEI Technologies · Google · HITACHI Inspire the Next · IBM · iMerit · inspur · intel REALSENSE TECHNOLOGY · NSA WHERE INTELLIGENCE GOES TO WORK IntelligenceCareers.gov/NSA · NVIDIA · OAK RIDGE National Laboratory · Uber · WAYMO

cvpr20.com

IEEE COMPUTER SOCIETY · CVF

COMPUTING edge

IEEE COMPUTER SOCIETY computer.org

IEEE COMPUTER SOCIETY | IEEE | Printed with inks containing soy and/or vegetable oils

STAFF

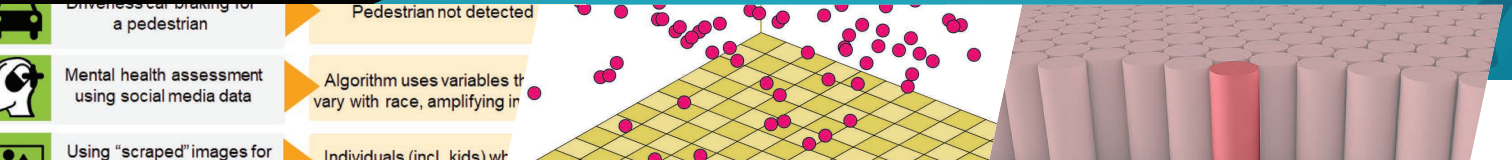
Circulation: *ComputingEdge* (ISSN 2469-7087) is published monthly by the IEEE Computer Society. IEEE Headquarters, Three Park Avenue, 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720; voice +1 714 821 8380; fax +1 714 821 4010; IEEE Computer Society Headquarters, 2001 L Street NW, Suite 700, Washington, DC 20036.

Postmaster: Send address changes to *ComputingEdge*-IEEE Membership Processing Dept., 445 Hoes Lane, Piscataway, NJ 08855. Periodicals Postage Paid at New York, New York, and at additional mailing offices. Printed in USA.

IEEE Computer Society Magazine Editors in Chief

OCTOBER 2020 · VOLUME 6 · NUMBER 10

COMPUTING edge



Security and Privacy

Data

Internet

Artificial Intelligence

Departments

Magazine Roundup

The IEEE Computer Society's lineup of 12 peer-reviewed technical magazines covers cutting-edge topics ranging from software design and computer graphics to Internet computing and security, from scientific applications and machine intelligence to visualization and microchip design. Here are highlights from recent issues.

Computer

Reliability Inversion: A Cautionary Tale

Reliability analysis is often based on worst-case assumptions to produce guaranteed lower bounds on system survival probability. Reliability engineers make lower bounds as tight as possible, but sometimes system structure is unfriendly to the derivation of tight bounds. Unfortunately, loose reliability lower bounds make it difficult to compare design alternatives or to select among competing systems. Read more in this article from the June 2020 issue of *Computer*.

Computing in Science & Engineering

Design of a Virtual Reality Tour System for People With Intellectual and Developmental Disabilities: A Case Study

This article from the May/June 2020 issue of *Computing in Science & Engineering* focuses on VR as a form of therapy for individuals with intellectual and developmental disabilities (IDDs). The research aim is to develop an immersive and interactive VR system that is tailored for IDD individuals, for whom most currently available VR experience systems are not optimized. Being intimately familiar with a place through an interactive VR tour will help alleviate social anxiety. Accordingly, the authors create a hotspot-based VR tour system, which can provide an almost lifelike experience of visiting and learning about the location. They conducted experiments with nondisabled individuals, acting as the control group, and IDD individuals, acting as the experimental group, to evaluate the tour system and compare the results between two groups. The experiments show that the VR tour has a positive impact on the IDD individuals.

IEEE Annals of the History of Computing

The First Computer in New Zealand

How quickly did the computer revolution reach the most remote Westernized country? Conventional history holds that the first modern computer in New Zealand—where "modern" means electronic, and with stored programs—was an IBM 650 leased from IBM Australia by the New Zealand Treasury in November 1960, and officially inaugurated in March 1961. This article from the April–June 2020 issue of *IEEE Annals of the History of Computing* discusses an alternative hypothesis—that the pioneer was in fact an ICT 1201 ordered in 1959 and installed by the New Zealand Department of Education a few months before the arrival of the IBM 650.

IEEE Computer Graphics and Applications

Many Views Are Not Enough: Designing for Synoptic Insights in Cultural Collections

Cultural object collections attract and delight spectators. Yet, they also easily overwhelm visitors due to their perceptual richness and associated information. Similarly, digitized collections appear as complex, multifaceted phenomena, which can be challenging to grasp and navigate. Though visualizations can create various types of collection overviews, they do not easily assemble into a "big picture" or lead to an integrated understanding. The authors of this

article from the May/June 2020 issue of *IEEE Computer Graphics and Applications* introduce coherence techniques to maximize connections between multiple views and apply them to the prototype PolyCube system of collection visualization. With map, set, and network visualizations, it makes spatial, categorical, and relational collection aspects visible. For the essential temporal dimension, it offers four views: superimposition, animation, juxtaposition, and space-time cube representations. A user study confirmed that better-integrated visualizations support synoptic, cross-dimensional insights.

IEEE Intelligent Systems

A Deep Coupled LSTM Approach for USD/CNY Exchange Rate Forecasting

Forecasting CNY exchange rate accurately is a challenging task due to its complex coupling nature, which includes market-level coupling from interactions with multiple financial markets, macro-level coupling from interactions with economic fundamentals, and deep coupling from interactions of the two aforementioned kinds of couplings. This article from the March/April 2020 issue of *IEEE Intelligent Systems* develops a new deep-coupled long short-term memory (LSTM) approach to capture the complex couplings for USD/CNY exchange rate forecasting. In this approach, a deep structure consisting of stacked LSTMs is built to model the complex couplings. The experimental results with 10 years of data indicate that the proposed approach significantly outperforms seven other benchmarks. The purpose in this article is to clarify the importance of coupling learning for exchange rate forecasting, and the usefulness of the deep-coupled model to capture the couplings.

IEEE Internet Computing

Archipelago: A Medical Distributed Storage System for Interconnected Health

With the development of the Internet and personal wearable devices, people pay more attention to the storage and application of wearable device data, which also brings more challenges. Personal health data stored in the clouds has the risk of privacy leakage. Also, health data of individuals, community clinics, and hospitals are all information islands. In this article from the March/April 2020 issue of *IEEE Internet Computing*, the authors design a system named Archipelago, which makes it possible to connect individuals, community clinics (or family doctors), and hospitals to share information. The authors use the method based on system information perception to segment data and decide the location of block location, making full use of nodes with different storage capacity and performance.

IEEE micro

Energy-Efficient Video Processing for Virtual Reality

Virtual reality (VR) has huge potential to enable radically new applications, behind which spherical panoramic video processing is one of the backbone techniques. However, current VR systems reuse the techniques designed for processing conventional planar videos, resulting in significant energy inefficiencies. The authors of this article from the May/June 2020 issue of *IEEE Micro* present EVR, an end-to-end system for energy-efficient VR video processing. EVR recognizes that the major contributor to the VR tax is the projective transformation (PT) operations. EVR mitigates the overhead of PT through two key techniques: semantic-aware streaming on the server and hardware-accelerated rendering on the client device. Real system measurements show that EVR reduces the energy of VR rendering by up to 58%, which translates to up to 42% energy saving for VR devices.

IEEE MultiMedia

Building a Manga Dataset "Manga109" With Annotations for Multimedia Applications

Manga, or comics, which are a type of multimodal artwork, have been left behind in the recent trend of deep-learning applications because of the lack of a proper dataset. The authors of this article from the April–June 2020 issue of *IEEE MultiMedia* built Manga109, a dataset consisting of 109 Japanese comic books (94 authors and 21,142 pages), and made it publicly available by obtaining author permissions for academic use. They carefully annotated the frames, speech texts, character faces, and character bodies; the total number of annotations exceeds 500k. This dataset provides numerous manga images and annotations, which will be beneficial for use in machine-learning algorithms and their evaluation. In addition to academic use, the authors obtained further permission for a subset of the dataset for industrial use.

IEEE Pervasive Computing

Betrusted: Improving Security Through Physical Partitioning

The condensation of virtually everything into a single device—the smartphone—has normalized deviant behaviors that create security risks. For example, many smartphone users conduct secure transactions while juggling several other apps, thus creating opportunities for adversaries to exploit human error. Furthermore, both secure and insecure code run on a smartphone's common CPU, thus exposing user secrets to a large and complex attack surface with multiple microarchitectural side channels. This article from the April–June 2020 issue of *IEEE Pervasive Computing* proposes partitioning a set of secure applications into a physically separate device that is designed using security-first principles.

IEEE Security & Privacy

Benchmarking Flaws Undermine Security Research

Benchmarking systems is difficult. Mistakes can compromise guarantees and threaten reproducibility and comparability. The authors of this article from the May/June 2020 issue of *IEEE Security & Privacy* conduct a study to show that benchmarking flaws are widespread in systems security defense papers, even at tier-1 venues. The authors aim to raise awareness and provide recommendations for safeguarding the scientific process in our community.

IEEE Software

Metamorphic Testing: Testing the Untestable

What if we could know that a program is buggy, even if we could not tell whether its observed output is correct? Metamorphic testing provides this ability. This article from the May/June 2020 issue of *IEEE Software* explains the basics of the technique.

IT Professional

AI Down on the Farm

Agriculture has become an information-intensive industry. In the production of crops and animals, precision agriculture approaches have resulted in the collection of spatially and temporally dense datasets by farmers and agricultural researchers. These big datasets, often characterized by extensive nonlinearities and interactions, are often best analyzed using machine learning (ML) or other artificial intelligence (AI) approaches. In this article from the May/June 2020 issue of *IT Professional*, the authors review several case studies where ML was used to model aspects of agricultural production systems and provide information useful for farm-level management decisions. Taken together, these examples represent the current abilities and future potential for AI applications in agricultural production systems.

Editor's Note

Cybersecurity and COVID-19

One of the many secondary effects of the COVID-19 pandemic is increased use of the Internet, videoconferencing, and other technologies people use to work remotely and communicate while social distancing. In some cases, this increased demand has strained networks and exposed previously unknown vulnerabilities. In addition, cybercriminals are exploiting the situation with new strategies. This issue of *ComputingEdge* explores how the pandemic is affecting computer security and privacy, as well as how to mitigate risks and attacks.

IEEE Security & Privacy's "Isolating Insecurely: A Call to Arms for the Security and Privacy Community During the Time of COVID-19" urges cybersecurity professionals to contribute to the fight against emerging threats related to the pandemic. *IT Professional*'s "IT Risk and Resilience—Cybersecurity Response to COVID-19" examines the pandemic's effects on information technology. The authors recommend that organizations plan for crises and take steps to prevent and deal with cyberattacks.

Data security and privacy were concerns long before the pandemic hit. The authors of *IEEE MultiMedia*'s "Legal and Ethical Challenges in Multimedia Research" argue for ensuring that data collected through multimedia research—such as traffic video and images of people—is kept private and used with consent. *IEEE Software*'s "Bringing Semantic Knowledge Graph Technology to Your Data" describes a semantic search and classification system that organizations and individuals can utilize to manage their private data.

One popular data-gathering technique is online crowdsourcing. In *IEEE Micro*'s "Expertise at Our Fingertips," the author analyzes the world's most influential Internet crowdsourcing project: Wikipedia. *IEEE Internet Computing*'s "Teaching Crowdsourcing: An Experience Report" describes an engaging assignment in which college students design a crowdsourcing project for hummed song recognition.

This *ComputingEdge* issue closes with *Computer*'s "Biologically Driven Artificial Intelligence." The authors posit that leveraging computational principles of the human brain, rather than artificial neural networks, will make AI smarter.

This article originally appeared in
IEEE SECURITY & PRIVACY
vol. 18, no. 4, 2020

Isolating Insecurely: A Call to Arms for the Security and Privacy Community During the Time of COVID-19

Sean Peisert, *Associate Editor in Chief*

Let's begin by saying that this was not the column I had originally planned to write. However, I have the dubious distinction of composing this piece while the world is dealing with the coronavirus (COVID-19) pandemic. As I type this, substantially more than 1 million confirmed cases have been reported, and tens of thousands of people have died from the disease. Surely, those figures will have risen considerably by the time this column is published. Very thankfully, hundreds of thousands more who were positively diagnosed have now recovered.

Undeniably, our first duty is to keep ourselves, our families, our communities, and the general public around the world safe by following shelter-in-place rules and related public health measures to contain the outbreak. Beyond that key measure, what else can the rest of us do to help improve the situation, perhaps beyond donating compute cycles on our laptops to the Folding@home COVID-19 effort?[1]

I find myself observing many colleagues in the life sciences participating directly in the response effort. Government agencies and philanthropists are sponsoring massive efforts to track the spread of the virus, set up testing facilities, manufacture protective equipment, analyze the virus and its mutations, develop vaccines, and more. For the rest of us without advanced degrees in domains such as virology, microbiology, biochemistry, and biomedicine, being unable to directly contribute to the response can instill a helpless feeling, and I admit to experiencing some envy toward my life sciences colleagues who are helping to prevent further casualties around the world.

At the same time, all of us can observe, unfortunately, that attackers have not taken a break from their usual activities. In many cases, their assaults may simply reflect run-of-the-mill escalations of the usual activity on the Internet—through an increase in COVID-19-related phishing,[2] for example—to take advantage of the increased anxiety and distraction resulting from the situation. Other attacks are more specific to the situation, more targeted, and more destructive. It is here where, even though as cybersecurity and privacy professionals we are not on the front lines of the virus response, we must rise to the occasion to provide support for the individuals and institutions who are on those front lines and who will ultimately bring us through and past this situation.

Much of planet Earth is now undergoing some form of sheltering in place. In addition, many of those who are still employed are teleworking. For those of us who live in areas with reasonable broadband network access, we are thankful that the Internet can help get us through this time, whether by video socializing, ordering groceries online, or attending work or school. Indeed, around the world, students from kindergarten all the way through graduate school are trying to participate in distance-learning activities. Recently, the lynchpin for much of this remote school and work activity, other than email and basic broadband Internet connectivity (which is not available in many rural areas across the globe), has been videoconferencing, particularly the service provided by Zoom Video Communications. In many cases, Zoom is now a lifeline that the world has

Digital Object Identifier 10.1109/MSEC.2020.2992316
Date of current version: 9 July 2020

 Published by the IEEE Computer Society

suddenly come to depend on to be able to function in an environment of physical isolation.

Alas, let us count the ways in which Zoom has failed its customers, with respect to security and privacy.[3] Zoom's security-related processes and major safety flaws will undoubtedly be rectified, and its privacy practices will be addressed, now that the company has been put in the spotlight. However, Zoom's shortcomings are only one part of the story. The other part concerns the huge and rapid adoption of a software service and the nature of that application's new user base. Zoom's traditional market has consisted of business customers, which often have IT departments and security teams. In contrast, given the COVID-19 situation, tens of millions of very inexperienced people have begun to use Zoom for online classes and corporate meetings, and even some very large public assemblies, and they often administer and utilize the service without sufficient expertise or the training typically required for such an application. The consequences of inexperienced users adopting key software are something that all developers should keep in mind. It has been a fundamental tenet of software engineering for decades that users should never be required to do what developers might expect.

While "Zoombombing" exemplifies the perfect storm of software flaws meeting inexperienced users, it is representative of a much broader set of computer security problems created by the pandemic. The more worrisome examples that come to mind include the researchers working on COVID-19 who are now storing regulated data on their home computer systems. There are testing labs running networked, automated machinery to examine thousands of samples per day to detect the presence of the virus only to fall victim to ransomware.[4] Imagine the potential effects of malware disrupting testing labs at scale, reducing or even tainting the facilities' output; experts in cyberphysical system security are immediately needed to examine the processes and build in the necessary safeguards. Think about the global surveillance networks that are looked upon to provide authoritative information about the extent of the virus's spread; they are also highly vulnerable.[5] Consider the potential effects of tainting the inputs to the global surveillance network; expertise in data integrity, fault tolerance, cryptography, and other related disciplines is clearly necessary to make the network more resilient. In all these cases and in related areas, we as cybersecurity and privacy professionals realize where we can best devote our attention at a time when society is at its most reliant on vulnerable digital systems.

I was heartened to read of the COVID-19 Cyberthreat Intelligence (CI) League of security professionals, which was formed as a public service to help mitigate global cybersecurity threats, among others.[6] The scientific-computing community, including one organization with which I am involved, the National Science Foundation Cybersecurity Center of Excellence, Trusted CI, has also stepped up to the plate to help support cybersecurity needs during this time.[7] Thus, like looters after the London air raids or Atlantic hurricanes, it is discouraging, though not surprising, to see miscreants mobilize for their own benefit. It is at least as encouraging to see the other side of the coin, where the public, including cybersecurity professionals, has volunteered to support the common good.

One note of caution is that volunteer work from the security and privacy community must primarily or even exclusively be in the service of the cause, not in the name of future fame and glory from research publications. Security and privacy researchers descending on medical professionals to gather data for their next conference publication are likely doing more harm than good. But researchers who are genuinely interested in bringing the right tools to bear on solving the problem at hand by listening to medical professionals and putting clinicians' needs ahead of all others, as I wrote about in this space last year, are desperately needed.

There is much more work to do. The privacy question is one that has only begun to be addressed. While many of the staunchest privacy advocates have argued for the relaxation of privacy controls to develop better tests, back-to-work protocols, and vaccines and other treatments, the same activists point out that any loss of privacy for the public good should be temporary, transparent, necessary, and proportionate and that it should follow a due process.[8,9] What should solutions look like for data gathering, sharing, use, and disposal? The same question applies protocols requiring strong individual identity verification and validation. Nobody really knows the answers yet.

Limited waivers of enforcement sanctions and penalties under the Health Insurance Portability and Accountability Act privacy rules in the United States have already been put into place.[10] What direction will European countries take with respect to the General Data Protection Regulation? Will COVID-19 mitigation needs lead to the public's electronic health records and gene sequences being made even more broadly available for analysis, perhaps dramatically so? If that happens, how will it be done, and what will the effect be? Will the United States and European countries adopt somewhat Orwellian-sounding smartphone-based all-clear and free-to-travel indicators, as China has, or will a solution come to light that more strongly preserves individual privacy? There is a need and opportunity for experts in privacy-preserving analysis computation techniques, such as differential privacy and secure multiparty computation, to very quickly engage with stakeholders to bring usable and practical solutions to bear on these critical problems while properly balancing key privacy and analysis properties.

Meanwhile, in the United States and elsewhere throughout the world (e.g., Croatia, Egypt, Iceland, New Zealand, and Poland), all of this is happening in an election year, during which many primaries and perhaps even general elections may move to remote-voting processes. In some cases, the shift will be to a vote-by-mail program. For large states that still conduct a majority of voting in person, the transition might be challenging: securely managing lists of eligible voters, printing and mailing ballots, employing automated signature-comparison software, and safely leveraging systems to automatically count paper ballots. However, we now see numerous U.S. states advocating for a shift to Internet voting, which seems ripe for disaster, given observations of past attempts in this area, the unequivocal conclusions of a recent National Academies report ("…the Internet should not be used for the return of marked ballots. … [N]o known technology guarantees the secrecy, security, and verifiability of a marked ballot transmitted over the Internet."[11,12]), and a letter from the American Association for the Advancement of Science's Center for Scientific Evidence in Public Issues (and other experts).[13] Legislatures of various sizes and across a range of jurisdictions are also considering online voting.

Even without the scale of U.S. public elections and the typical requirement for preventing the association of individual voters with cast ballots, legislative voting presents its own set of challenges that must be carefully resolved to ensure proper security.[12] The importance of having election and computer security experts help policy makers and election officials understand the conclusions from the National Academies report—even or perhaps especially in light of the pandemic—cannot be understated. Readers of this piece: get engaged in anything from volunteer support for your city and county elections IT staff to helping to educate policy makers.

Thankfully, again, I am heartened to see that the cybersecurity community seems prepared to step up to face these challenges. In addition to the aforementioned community efforts, Apple and Google announced privacy-preserving solutions for contract tracing[14] (in addition to a variety of similar academic efforts[15,16]), largely with the approval of privacy experts, and the companies have deployed differential privacy and cryptographic mechanisms to monitor mobility without exposing individual identities.[17] I also see academic colleagues with expertise in machine learning addressing active disinformation campaigns that are spreading conspiracy theories about causes of and remedies for COVID-19.[18] Finally, I observe polling-process experts helping to educate policy makers, voting officials, and the public about the best paths forward for conducting elections by remote means (by mail, using paper ballots).

While I envy those who can directly contribute to the biological or medical portion of the crisis at hand, I'm also glad to support their response through security and privacy as well as by providing the infrastructure that we are relying on to get us through this time and that will eventually have a key role in putting society back together again. For my part, it's an all-out effort to contribute everything I have to the cause. For those security and privacy professionals who can, please join me. The challenges we face now and during the coming months require our expertise. This is the time to devote your energy to being part of the home guard, to keeping things running while the virologists and geneticists develop the solution that stems the tide of layoffs, infection, and

death—the resolution that enables people to take off their masks and go back to work to earn livings and that allows the children return to school and the playgrounds to reopen.

REFERENCES

1. "*COVID-19*," Folding@home. Accessed on: Apr. 14, 2020. [Online]. Available: https://foldingathome.org/covid19/
2. "*Defending against COVID-19 cyber scams*," U.S. CERT, Washington, D.C., Mar. 6, 2020. [Online]. Available: https://www.us-cert.gov/ncas/current-activity/2020/03/06/defending-against-covid-19-cyber-scams
3. G. Fleishman, "Every Zoom security and privacy flaw so far, and what you can do to protect yourself," *TidBITS*, Apr. 3, 2020. [Online]. Available: https://tidbits.com/2020/04/03/every-zoom-security-and-privacy-flaw-so-far-and-what-you-can-do-to-protect-yourself/
4. I. Ilascu, "*COVID-19 testing center hit by cyberattack*," BleepingComputer, Mar. 14, 2020. [Online]. Available: https://www.bleepingcomputer.com/news/security/covid-19-testing-center-hit-by-cyberattack/
5. B. Schneier, "*Security of health information*," Schneier on Security, Mar. 5, 2020. [Online]. Available: https://www.schneier.com/blog/archives/2020/03/security_of_hea.html
6. "International cybersecurity experts come together to fight COVID-19 related cyberthreats," *CISOMAG*, Mar. 31, 2020. [Online]. Available: https://www.cisomag.com/international-cybersecurity-experts-come-together-to-fight-covid-19-related-cyberthreats/
7. V. Welch, "*Trusted CI, NSF CI CoE Pilot, and SGCI offering priority help to projects tackling COVID-19*," Trusted CI Blog, Mar. 17, 2020. [Online]. Available: https://blog.trustedci.org/2020/03/trusted-ci-nsf-ci-coe-pilot-and-sgci.html
8. C. Cohn, "*EFF and COVID-19: Protecting openness, security, and civil liberties*, Mar. 23, 2020. [Online]. Available: https://www.eff.org/deeplinks/2020/03/eff-and-covid-19-protecting-openness-security-and-civil-liberties
9. New York Times Editorial Board, "Privacy cannot be a casualty of the coronavirus," *NY Times*, Apr. 7, 2020. [Online]. Available: https://www.nytimes.com/2020/04/07/opinion/digital-privacy-coronavirus.html
10. "*COVID-19 & HIPAA Bulletin: Limited waiver of HIPAA sanctions and penalties during a nationwide public health emergency*," U.S. Dept. of Health and Human Services, Washington, D.C., Mar. 2020. [Online]. Available: https://www.hhs.gov/sites/default/files/hipaa-and-covid-19-limited-hipaa-waiver-bulletin-508.pdf
11. National Academies of Sciences, Engineering, and Medicine, *Securing the Vote: Protecting American Democracy*, Washington, D.C.: The National Academies Press, 2018.
12. A. Appel, "*Can legislatures safely vote by Internet?*" Freedom to Tinker, Apr. 10, 2020. [Online]. Available: https://freedom-to-tinker.com/2020/04/10/can-legislatures-safely-vote-by-internet/
13. "*Letter to Governors and Secretaries of State on the insecurity of online voting: Letter from AAAS EPI Center and leading experts in cybersecurity and computing*," American Association for the Advancement of Science, Washington, D.C., Apr. 9, 2020. [Online]. Available: https://www.aaas.org/programs/epi-center/internet-voting-letter
14. "*Apple and Google partner on COVID-19 contact tracing technology*," Apple Inc., Cupertino, CA, Apr. 10, 2020. [Online]. Available: https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/
15. C. Troncoso et al., "*Decentralized privacy-preserving proximity tracing*," Decentralized Privacy-Preserving Proximity Tracing (DP-3T) project, Apr. 10, 2020. [Online]. Available: https://github.com/DP-3T/documents
16. R. Rivest et al., "*PACT: An open, privacy-preserving protocol*," PACT: Private Automated Contact Tracing, Apr. 8, 2020. [Online]. Available: https://pact.mit.edu/wp-content/uploads/2020/04/The-PACT-protocol-specification-ver-0.1.pdf
17. "*Mobility trends reports*," Apple, Cupertino, CA. [Online]. Available: https://www.apple.com/covid19/mobility
18. J. E. Barnes, M. Rosenberg, and E. Wong, "As virus spreads, China and Russia see openings for disinformation," *NY Times*, Mar. 28, 2020. [Online]. Available: https://www.nytimes.com/2020/03/28/us/politics/china-russia-coronavirus-disinformation.html

SEAN PEISERT, Associate Editor in Chief

This article originally appeared in
IT Professional
vol. 22, no. 3, 2020

IT Risk and Resilience—Cybersecurity Response to COVID-19

Tim Weil, *SecurityFeeds LLC*
San Murugesan, *Western Sydney University*

The rapid and worldwide spread of the coronavirus and its illness known as COVID-19 has made huge impact on almost everything has taken us all by surprise. We all are now experiencing a major unprecedented and unexpected global public health crisis. This pandemic has also triggered huge social upheavals, disrupted almost every industry, and impacted the life and work of everyone in almost every country. Businesses and educational institutions are closed, many employees are forced to work from their homes, supply chains have been disturbed, people are being required to self-isolate, and most travel, in-person meetings, and conventions have been banned. These disruptions could continue for months, and the resulting economic, business, and social impact will last for years.

Nevertheless, business operations and services must continue on, effectively and uninterrupted. IT has been employed in novel and traditional ways to meet these challenges. Migration of many operations and services online for remote work has become inevitable, and technologies, such as cloud computing, robots, drones, AI, chatbots, VPN, virtual dashboards, autonomous systems, and the Internet facilitate this digital transformation. IT has now taken a central role in every activity and has become an epicenter of operations in healthcare, business, education, governance, judiciary, community service, and more. What and how we do our daily personal and business activities are significantly transformed with the aid of recent developments in IT, as outlined in Table 1. It is very likely that even after we successfully emerge from the crisis, business will not be "as usual" and we may continue new ways of working and offering various services.

The COVID-19 epidemic impacted IT too, primarily positively, benefiting IT industry and IT professionals and serving public goods. However, there are a few negative impacts as well, such as increased and novel cybersecurity threats and risks, performance issues due to significantly increased workload, and business continuity (BC), which the IT industry has tackled satisfactorily.

In this context, we, IT professionals and business executives, have to critically examine the following key questions:

- Are the IT industry and other enterprises prepared for this makeover or change, and how well?
- How has the IT industry responded to explosion in demand for traditional and newer services? What innovations has this epidemic brought about? What else can we do?
- Did this crisis expose cracks in our current IT planning and offerings, and business/IT risk management? What are they?
- What has been the impact on its performance of significant increase in widespread use of IT?
- What are the security and other risks the new operational environment poses and how can we assess and address them?
- What lessons can we learn from responding to this crisis?

Digital Object Identifier 10.1109/MITP.2020.2988330
Date of current version 21 May 2020.

 Published by the IEEE Computer Society

TABLE 1. Global transformation caused by the coronavirus.

Industry	Response/Impact	Response	Underlying technology/ operation
Education	Widespread closure of educational institutions; access to labs is restricted; projects have been mothballed; and fieldwork interrupted	Virtual learning environment (online teaching, presentation, assessment, and consultation); convocation online	Online video conferencing software, virtual labs on cloud
Healthcare	Overcrowded hospitals, inability to meet the demands on them	Contact tracing, forecasting resource requirements, allotment of scare resources based on a patient's survivability, COVID-19 vaccine development, telehealth (online consultation with a doctor or medical professional); automated diagnosis	AI, ML, cloud computing, chatbot
Business	Closure of business, avoidance ofin-person retail shopping	Adherence to social distancing, services online, work from home	Chatbot, drone delivery, online meeting software, virtual office/desktop, remote access to work
Industry	Closure of business, avoidance of in-person retail shopping	Work from home, remote operations, automation and autonomous operation	Robots, automation, 3-D printing
Retail	Stores closed, only online service, avoidance of retail shopping	Online shopping, home delivery	The Web, online payment, contactless payment
Government	Spike in demands from citizens for assistance, disruption to normal operations	Migration to online services	Cloud, the Web, online meeting application
Entertainment	Entertainment venues (parks, cinema) closed, sports without spectators	Viewing online	Audio and video streaming, virtual reality
Personal life and social interaction	Lockdown	Indoor activities	Phone, audio and video chats, streaming, online gaming
Spirituality and religious practices	Places of worship closed	Online participation, prayers from home, worship through livestream	Audio and video streaming, virtual reality
Conferences	In-person conferences banned; virtual conferences	Online presentation and discussion	Video streaming, virtual conference software

- How will COVID-19 reshape IT, IT security, and risk assessment and management?
- How can we proactively plan to successfully handle crises that we might face in the future?

By addressing these critical questions—not only during the COVID-19 crisis, but also regularly—as a standard practice, we will be better prepared for whatever comes. In this article, we examine some of these questions. We also invite you to share your thoughts and ideas.

FIGURE 1. NIST Cybersecurity Framework.

IT SECURITY DURING THE PANDEMIC CRISIS

Pandemic events stress test IT systems, tactical security measures, and IT governance models causing strategic (long-term) disruption in the global digital fabric. The cybersecurity impact of the COVID-19 pandemic has spread to all sectors of international commerce including citizens, industry, government, and academic sectors. Cybersecurity professionals

TABLE 2. Threats and vulnerabilities caused by the coronavirus.

Threats and vulnerabilities	Online resource
ZOOM bombing	https://delta.ncsu.edu/news/2020/04/02/zoom-security-and-privacy/
Spyware and phishing	https://www.coalfire.com/The-Coalfire-Blog/March-2020/COVID-19-incites-cyber-crimes-of-opportunity
Malware and phishing	https://www.webarxsecurity.com/covid-19-cyber-attacks/
Health check—ISPs, cloud providers, UCaaS during pandemic	https://www.networkworld.com/article/3534130/covid-19-weekly-health-check-of-isps-cloud-providers-and-conferencing-services.html

are urgently responding to increased cyber threats and their responses span the spectrum of information security and privacy management capabilities.

The NIST cybersecurity framework (CSF) which consists of Identify, Protect, Detect, Respond, and Recover functions (see Figure 1) offers a lightweight model for companies to address the new threats and attack surface presented by COVID-19 cybersecurity earthquake.[1] Details of the framework and how diverse organizations used the methodology to improve their cybersecurity risk management are provided on the NIST CSF portal.[2] We use the CSF model to frame our discussion of global cybersecurity response. Our story highlights a set of tables showing the CSF method and industry response examples to illustrate that "there is a method to the madness" of our cybersecurity response to COVID-19.

IMPACT ON GLOBAL IT

Across the board, the cybersecurity industry has identified major threats, vulnerabilities, and attack vectors and responded with recommendations for risk management, continuity planning, containment, remediation, and recovery solutions. Sudden and massive migration to remote work has required business entities to equip and enable their IT systems for remote work and manage personnel in unprecedented new ways. The NIST CSF IDENTIFY function assists in developing an organizational understanding to managing cybersecurity risk to systems, people, assets, data, and capabilities.[1]

INCREASED THREATS AND VULNERABILITIES

The NIST CSF Protect function outlines appropriate safeguards to ensure delivery of critical infrastructure services. The Protect function supports the ability to limit or contain the impact of a potential cybersecurity event.[3] Leveraging the COVID-19 anxiety and concerns and the absence of on-site personnel support, new attack vectors have emerged and gotten significant coverage in the media and the cybersecurity industry. Table 2 provides industry examples of the CSF PROTECT/DETECT activity. They include the following.

ZOOM Bombing—Security and privacy vulnerabilities in teleconferencing software allow trolling hackers to intercept authentication credentials and inject objectionable content (such as pornographic materials and violent images) into seemingly secure collaborative online meetings. An example of a ZOOM bombing exploit is interference with academic networks such as a thesis defense given over a university teleconference.

COVID-19 Phishing Attacks—As reported in FBI bulletins, there were fake, malicious emails that appeared to be from the Center for Disease Control (CDC). They contained malware attachments, or aimed to hijack user credentials.

Malware—An example of malware is a Corona Trojan overwriting master boot record and disabling hard disk storage. Ransomware attacks on healthcare systems have been escalating during the pandemic. Table 3 illustrates snapshots of malware and phishing attacks.

Network Availability—While performance of core communication networks and clouds remained satisfactory despite substantial increase in traffic, some collaborative applications faced spikes in service outages, as shown by the Network World example in Table 2.

The NIST CSF Detect function defines the appropriate activities to identify the occurrence of a cybersecurity event. The Detect function enables timely discovery of cybersecurity events as shown in Table 3.

TABLE 3. Sample of malware and phishing attacks reported during coronavirus crisis.

Date	Description of cyberattack	Type of attacks
April 8, 2020	The exposure to compromised e-commerce websites is greater than ever. 26% increase in web skimming in March.	Malware
April 8, 2020	"Latest vaccine release for coronavirus (COVID-19)" mall spam spreads NanocoreRAT malware	Malware
April 8, 2020	NCSC advisory: COVID-19 exploited by malicious cyber actors	Social engineering
April 8, 2020	Fake COVID19 website is spreading FirebirdRAT via fake DHL emails	Malware
April 8, 2020	Rush to adopt online learning under COVID-19 exposes schools to cyberattacks	Zoom bombing
April 8, 2020	Sophisticated COVID-19-based phishing attacks leverage PDF attachments and SaaS to bypass defenses	Phishing, malware
April 8, 2020	CDC warns of COVID-19-related phone scams, phishing attacks	Phishing

Source: (Webarxsecurity Website, https://www.webarxsecurity.com.)

TABLE 4. Cybersecurity risk mitigation and response to the coronavirus.

Cybersecurity management response	Online resource
CxO Education (security architects partners)	https://security-architect.com/waking-up-to-the-new-covid-19-cybersecurity-reality/
COVID-19 Joint Acquisition Task Force	https://www.acq.osd.mil/jatf.html
US DHS Cyber and Infrastructure Agency (CISA)	http://www.cisa.govsites/default/files/publications/20_0306_cisa_insights_risk_management_for_novel_coronavirus.pdf
NIST SP 800-46 Guide to enterprise telework, remote access, and BYOD security	https://csrc.nist.gov/publications/detail/sp/800-46/rev-2/final

INDUSTRY RESPONSE

CEOs are often asked "what keeps you awake at night"? In response to the new COVID-19 threats, the CEO answer may well be "everything!" Strategic consulting firm, Security Architect Partners (SAP) recommends clear dialogues at the CxO management level, weaving key COVID-19 cybersecurity issues like employee morale, BC, telecommuting, or supply chain management into normal executive meetings. SAP recommendations to address top security concerns in COVID-19 times include:

- securing remote access;
- mitigating increased fraud and malware threats;
- assessing new suppliers' (and changes to existing supplier's security posture);
- protecting core information system availability and security;
- refactoring security program priorities, architectures, and budgets;
- managing work force morale; and
- aligning with business leadership.

The NIST CSF Respond function includes appropriate activities to take action regarding a detected cybersecurity incident. The Respond function supports the ability to contain the impact of a potential cybersecurity incident.[3] In the US, at the federal agency level, the use of the NIST CSF has long been incorporated into the cybersecurity management fabric of risk mitigation. As shown in Table 4, examples of agency response to the pandemic include the Department of Homeland Security (DHS) and Department of Defense information portals. Across all US government services, contingency plans have been activated and the disruption of service operations continues to

TABLE 5. Cybersecurity response to the coronavirus.

Cybersecurity risk mitigations	Online resource
Organizational resilience (Deloitte)	https://www2.deloitte.com/content/dam/Deloitte/global/Documents/About-Deloitte/CoronaVirus_POV_People%20Technology%20Path_Global_Final%20(002).pdf
Legacy Software (COBOL) Supporting Financial Systems	https://edition.cnn.com/2020/04/08/business/coronavirus-cobol-programmers-new-jersey-trnd/index.html
Fired Americans Send Unemployment Websites Crashing Down	https://www.bloomberg.com/news/articles/2020-03-25/fired-americans-send-state-unemployment-websites-crashing-down

be restored. In the area of remote access, NIST special publication 800-46 offer guidance on host security, information security; network security, remote access, bring your own device (BYOD) and telework.

RESILIENCE AND RECOVERY

By and large, IT services, and IT industries in almost every country are coping with the demand on them and addressing the challenges due to the COVID-19 crisis. Speed bumps ahead are expected. In conversations with Dan Blum, Executive Adviser and author of the upcoming book "Rational Cybersecurity for the Business," we discussed the challenges customers face when force-fitting VPN architectures into demanding use cases, such as privileged third party access to their hybrid-cloud infrastructures. In order to enable business applications to work exclusively through remote access, some layers of defense (provided through firewalls and network segmentation) need to be modified or removed. This is worrisome in light of multiple nation state and cybercrime actors targeting unpatched VPN systems.

BY AND LARGE, IT SERVICES, AND IT INDUSTRIES IN ALMOST EVERY COUNTRY ARE COPING WITH THE DEMAND ON THEM AND ADDRESSING THE CHALLENGES DUE TO THE COVID-19 CRISIS.

This process aligns with the NIST CSF Recover function, which identifies appropriate activities to maintain plans for resilience and to restore any capabilities or services that were impaired due to a cybersecurity incident. The recover function supports timely recovery to normal operations to reduce the impact from a cybersecurity incident.[1]

From what we have surveyed, there have not been major IT system/service failures despite huge unexpected demands on IT infrastructure/services. For example, the Internet did work well despite high demand on it; clouds were able to scale with demand; start-ups and major companies were able to develop and deploy novel applications to address unique needs to contain the virus spread. Even in cases, where demands on customer-facing government applications (in Australia) peaked several times, interruption was temporary; IT operational staff addressed that overnight. Software was updated to meet new requirements in a short time. For instance, in Australia, application software was updated to enable special payment to millions of people affected by the lockdown by direct payment to their bank account within a couple of days after the policy/stimulus announcement.

Organizations are migrating most of the services and operations that can be performed online/remotely. Consequently there are now many vulnerabilities affecting the remote work force (employees, third-party vendors, and home office networks). BYOD, unpatched routers, and open WiFi present wider targets for hackers and intruders.

In the context of Leavitt's System Model of People, Process, Technology and Structure, the global health crisis hits all the targets that cybersecurity programs are designed to protect. As one of many examples, the international consulting firm, Deloitte Global Technology, offers a reasonable scenario for dealing with the issues of seismic organizational change and effective BC strategy as noted in Table 5. Highlights of the Deloitte continuity strategy recommendations include the following.

- Response strategy:
  - review BC/disaster recovery plans;
  - establish a crisis management office;
  - develop a communications plan.
- Personnel management (health and safety):
  - enforce precautionary measures and revisit sick leave policies;
  - review/amend policies for remote work, including guidelines on travel;
  - plan for absenteeism.
- Continuity of operation:
  - rationalize technology projects and portfolios;
  - equip your connectivity, security, and infrastructure for new traffic and use patterns;
  - be ready for disruptions in your business and technology ecosystem.

Legacy Software Support (COBOL)

To respond to the spike in financial assistance relief to their citizens, several state government agencies in the US urgently needed COBOL programmers to update their 40-year-old software. This has shown once again the limitations of persistence of legacy software for critical infrastructure and a lack of software programmers who are able to maintain and support these applications. A CNN article (cited in Table 5) highlights the need for legacy software support as cited by governors across the country.

IT has now taken a central role in every activity and has become an epicenter of operations in healthcare, business, education, governance, judiciary, community service, and more.

On top of ventilators, face masks, and health care workers, you can now add COBOL programmers to the list of what several states urgently need as they battle the COVID-19 pandemic. In New Jersey, Gov. Phil Murphy has put out a call for volunteers who know how to code the decades-old computer programming language called COBOL, because many of the state's systems still run on older mainframes. In Kansas, Gov. Laura Kelly said the state's Departments of Labor was in the process of modernizing from COBOL but then the virus interfered. "So they're operating on really old stuff," she said. Connecticut has also admitted that it is struggling to process the large volume of unemployment claims with its "40-year-old system comprised of a COBOL mainframe and four other separate systems."

FURTHER READINGS

1. COVID-19: Your IEEE resources. [Online]. Available: https://spectrum.ieee.org/static/covid19-ieee-resources?
2. COVID-19 Through the Business Technology Lens. [Online]. Available: https://www.cutter.com/covid-19-through-business-technology-lens
3. COVID-19 and the New Leadership Agenda. [Online]. Available: https://www.bcg.com/featured-insights/coronavirus.aspx?
4. P. V. Kannan, "Customer service continuity and lessons learned during the COVID-19 pandemic," Apr. 11, 2020. [Online]. Available: https://www.linkedin.com/pulse/customer-service-continuity-lessons-learned-during-covid-19-pv-kannan/

IT HAS NOW TAKEN A CENTRAL ROLE IN EVERY ACTIVITY AND HAS BECOME AN EPICENTER OF OPERATIONS IN HEALTHCARE, BUSINESS, EDUCATION, GOVERNANCE, JUDICIARY, COMMUNITY SERVICE, AND MORE.

PLANNING FOR THE FUTURE

Infectious disease outbreaks and other forms of crisis—anticipated and unanticipated—are inevitable. However, their impact can be mitigated through better preparedness and more effective responses. History shows that changes that we adopted in a crisis are not always temporary—crises can fundamentally reshape not only our beliefs and behaviors,[4] but also business and industry in many ways. And, IT will play even more crucial role in the post-COVID era.

IT and other industries must continue to proactively plan, focus on research and development on key areas of practical relevance, and revisit and tailor their policies. They also need to revisit and amend necessary crisis management policies and IT and business risk management policies, strategies, and practices taking lessons from the current crisis.

An organization's ability to effectively respond to a disruption not only depends on how effective it was in the planning process, but also how effective it was with its preparation, trials, and the training of their staff, which is often neglected.

CONCLUSION

The COVID-19 pandemic is a wakeup call to all of us. The world, IT, and our life and work post-Corona, will not be the same. In the context of IT, the pandemic has offered opportunities; exposed weaknesses and vulnerabilities of our IT systems and IT planning and implementation; and presented us—the IT industry, professionals, and governments—a few challenges.

In this short article, we examined a few aspects of IT risks and resilience (see also the sidebar, "Further Reading"). There is lot to think about, explore, plan, strategize, and act. Share your thoughts and ideas (by sending an e-mail to the authors) and join the new IEEE Computer Society's Special Technical Community, IT in Practice, an online platform for sharing technical knowledge and professional experiences.

REFERENCES

1. "Five functions of the cybersecurity framework," NIST. Apr. 2018. [Online]. Available: https://www.nist.gov/cyberframework/online-learning/five-functions
2. "Cybersecurity framework," NIST. Apr. 2018. [Online]. Available: http://www.nist.gov/cyberframework
3. "CISA INSIGHTS: Risk Management for Novel Coronavirus (COVID-19)," CISA. Mar. 18, 2020. [Online]. Available: https://www.cisa.gov/sites/default/files/publications/20_0318_cisa_insights_coronavirus.pdf
4. M. Reeves, et al., "Sensing and shaping the post-COVID era," Boston Consulting Group, Apr. 3, 2020. [Online]. Available: https://www.bcg.com/publications/2020/8-ways-companies-can-shape-reality-post-covid-19.aspx

TIM WEIL is currently a Cybersecurity Professional with SecurityFeeds LLC, Denver, CO, USA. He is a Senior Member of the IEEE and the Editor for *IT Professional* magazine. He is an industry-certified security professional (CISSP/CCSP, CISA, PMP, ISO 27001 Auditor) and an experienced auditor of enterprise security systems (federal, commercial). Contact him at trweil@ieee.org; http://www.securityfeeds.com.

SAN MURUGESAN is currently the Director of BRITE Professional Services, Sydney, NSW, Australia, and an adjunct professor with Western Sydney University, Penrith, NSW, Australia. He is a former Editor-in-Chief of the *IT Professional* magazine. He is a coeditor of *Encyclopedia of Cloud Computing (Wiley 2016), Harnessing Green IT: Principles and Practices* (Wiley, 2012) and other books. He is a Member of the COMPSAC Standing Committee, a Fellow of the Australian Computer Society, and the Institution of Electronics and Telecommunication Engineers and a Golden Core member of IEEE Computer Society. Contact him at san@computer.org; http://bitly.com/sanprofile.

DEPARTMENT: VISIONS AND VIEWS

This article originally appeared in IEEE MultiMedia vol. 27, no. 2, 2020

Legal and Ethical Challenges in Multimedia Research

Vivek K. Singh, *Rutgers University*
Elisabeth André, *University of Augsburg*
Susanne Boll, *University of Oldenburg*
Mireille Hildebrandt, *Radboud University, Netherlands and Vrije Universiteit Brussels*
David A. Shamma, *FX Palo Alto Laboratory*

Multimedia research has long moved beyond laboratory experiments and is being rapidly deployed in real-life applications including advertisements, search, security, automated driving, and healthcare. Hence, the developed algorithms now have a direct impact on the individuals using the abovementioned services and the society as a whole. While there is a huge potential to benefit the society using such technologies, there is also an urgent need to identify the checks and balances to ensure that the impact of such technologies is ethical and positive. For instance, if the multimedia technologies are being used to detect and protect pedestrians from accidents by autonomous vehicles, then the pedestrian detection performance needs to be equitable across demographic descriptors, such as gender and race of the pedestrians. Similarly, while logs of driving behaviors are important in many applications, making such information available to corporate entities and third parties could raise important privacy challenges.

This position article aims to: first, increase the awareness of such concepts and existing legal constraints in the multimedia research community, second, initiate a discussion on community guidelines on how to conduct multimedia research in a lawful and ethical manner, and third, identify some important research directions to support a vision of lawful and ethical multimedia research.

Recent growth spurt in multimedia research has led to some exciting developments in terms of multimedia content understanding and search, self-driving cars, and medical analysis. At the same time, there have been reports questioning both the processes and the outcomes for the developed technologies. For instance, Metz questions the ethics of using public image data in YFCC100M and IBM's Diversity in Faces datasets for training face recognition algorithms.[1] Under EU data protection law, any use must have a specific purpose and be limited to that purpose, while also requiring a valid legal basis. This clearly also goes for publicly available data, including images. Similar concerns have been raised about the outcomes of the developed algorithms. For instance, Buolamwini & Gebru has reported that face detection algorithms work much more accurately for white men than dark-skinned women, raising the

Digital Object Identifier 10.1109/MMUL.2020.2994823
Date of current version 12 June 2020.

 Published by the IEEE Computer Society

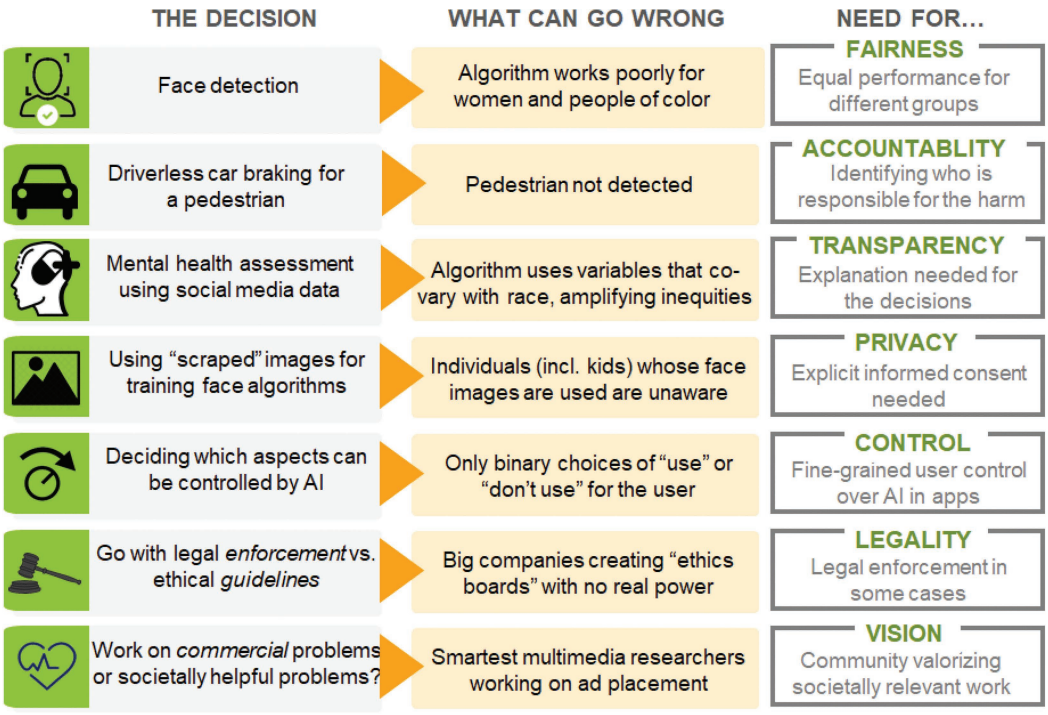


FIGURE 1. Summary of ethical challenges in multimedia research.

question whether dark-skinned women should require that their images become part of the training set or resist such inclusion (as it may be used for unwarranted surveillance purposes that have disparate effects for black women).[2] Further, multiple authors have criticized the use of video analysis software for automatic tracking of people in both civilian and military settings.[3] In fact, San Francisco has recently banned the use of face recognition technology for government applications.[4] Meanwhile, the Data Protection Authority in Hamburg, Germany has ordered Google to ban its employees "from listening to and reviewing EU data subjects" voice recordings for three months, to investigate potentially unlawful processing under the GDPR.[6]

Each of these issues raises important ethical concerns and many times the opinions of the experts in the multimedia community might not match with those in the popular media. There is an important need for the multimedia research community as a whole to have free and frank discussions on this topic and be cognizant of the myriad research and activism literature that is available regarding the potential benefits and harms of such technologies.[28–30] In doing so, the multimedia research community should not confuse ethical discussion with legal obligations as many of the so-called ethical concerns have clear answers (and obligations) as per law. In fact, a proper understanding of these obligations could lead the way to actionable respect for fundamental rights and freedoms at the level of research design. This will allow the multimedia research community to identify a set of community norms and guidelines on the processes and the outcomes of the technologies being developed. Developing such an understanding and a set of best practices would allow the multimedia research community to lead the conversation around these technologies rather than reacting to news stories about them (see Figure 1).

In this discussion, we must be careful not to confuse ethics with law. Many of the ethical challenges to be discussed below are part of the legal framework that applies to the multimedia applications based on machine learning. Since May 2018, the General Data Protection Regulation (GDPR) applies to the processing of personal data of people in the EU, whether or not the processing is done in the EU or by a company established in the EU. When relevant, we will discuss

the requirements of the GDPR, taking into account that applications meant to be deployed within the EU will have comply. The persistent confusion over what legal frameworks apply and what they mean for developers calls for dedicated attention to law for computer scientists; this viewpoint cannot do more than appetize the reader to take a deep dive into why and how law matters for their work.[31] The GDPR is one of the most advanced legislations, and though its scope is significant, many legal frameworks in other jurisdictions may have very different implications. The principles we highlight in this viewpoint are not necessarily anchored and enforceable at the global level. This means that whether and how they are legal or ethical principles is an empirical question. The fact that human rights courts have been weighing the corresponding fundamental rights against economic and public security interests for decades means that both ethics and computer science have lots to learn from the judicial scrutiny this has involved. Just like legislatures and courts have lots to learn from technical, scientific and ethical experts.

Fairness

Fairness means that the models developed do not systematically favor or disfavor a particular set of people. Angwin *et al.*[13] showed that parole decision algorithms being used in New York state were much more likely to assign positive outcomes to white defendants than black defendants. Buolamwini & Gebru found facial image based gender recognition algorithms to be much more accurate for white and male individuals than others.[2] Similarly, the dependence of multimedia algorithms for pedestrian detection on the age or race of pedestrians could result in unequivocally unfair outcomes.[25] Multimedia algorithms being used for parole decisions, driving decisions, and security applications can have important life-altering effects on people and it is important to ensure that the outcomes of the algorithms do not systematically favor or disfavor a specific set of people. Obviously, when algorithms detect that a particular characteristic of people correlates with higher risk (of recidivism, defaulting on loans, or causing road accidents), justice authorities, banks, or insurance companies will argue that treating these people differently is fair. An active research community at the intersection of machine learning, law and ethics is involved in this domain.[31–33] It is important that the multimedia community takes note of this and integrates the state-of-the-art design solutions to prevent violations of human rights, such as the right to nondiscrimination. In the research community dedicated to Fairness Accountability and Transparency (ACM-FAccT conferences), this has resulted in raising the more fundamental question of whether and when machine learning should be deployed, warning against computational solutionism.[33]

Accountability

"Algorithmic accountability ultimately refers to the assignment of responsibility for how an algorithm is created and its impact on society; if harm occurs, accountable systems include a mechanism for redress."[26] This is especially important in scenarios where there are multiple humans, companies, algorithms, and algorithm designers involved in the process.[7, 22] For instance, when a pedestrian is injured due to the decisions taken by a self-driving car, it is important to have accountability in place.[23] Going forward, if people are denied bail or organ transplants unfairly, it would be important to identify accountability in the process. However, liability obscurity associated with the use of modern ML algorithms is considered a major issue.[41] Attempts are being made towards certifying algorithms in order to enhance the transparency of accountability. It is, however, hard to predict whether multimedia systems that make use of highly dynamic ML algorithms will always behave according to a specification. Thus, certification is not a trivial task. Furthermore, it is unclear whether and how existing certification procedures can be adapted to modern multimedia systems. Part of a certification might include the use of state-of-the-art algorithms that check whether a multimedia system is suffering from bias. Finally, it is important to note that responsibility has to be taken over by humans and not by the machine, and usually there is not just one stakeholder in charge of it. The European Commission is currently considering adaptations of the liability regime for AI, making sure that accountability is not dependent on whatever a company may deem ethical, but on the need to compensate harm and damage.[34] This may result in more foresight and help developers propose the state-of-the-art applications that incorporate safety by design and

data protection by design. Here again, such accountability may result in the choice not to develop or use certain applications at all, as this may be the only responsible approach, considering the consequences.

Transparency

Algorithmic transparency is the principle that the factors that influence the decisions made by algorithms should be visible, or available, to the people who use, regulate, and are affected by systems that employ those algorithms.[17] Note that transparency and fairness are two different things. It is possible for a decision to be very transparent but not fair (e.g., admit only males) and vice-versa (e.g., a hypothetical neural network that ensures fairness but no one can understand why and how). The GDPR includes a "right to explanation" of decisions made by algorithms, whenever the decision is automated and has a significant effect on individual persons whose data are being processed.[35] In its rudimentary sense, many credit scoring applications in the US, Germany, and other countries identify the factors affecting a person's credit score, though given factors will often operate as a proxy for hidden variables that result in discrimination. Especially when deep learning algorithms have been used, a much higher level of transparency is required to figure out potential discrimination. This concerns behavioral targeting, where the EPIC (Electronic Privacy Information Center) has called for regulations that require advertisers to disclose the demographic factors behind targeted political ads, as well as the source and payment.[24]

Privacy and Data Governance

Multimedia research needs to ensure high quality of results in a way that also ensures human dignity in the process and in the results. This is often presented as a zero-sum game, but that is not necessarily the case.[36] Human subject research in medicine and the social sciences has a long-standing history of "informed consent" from the participants. While the web-based data collection is great for scaling up the studies, there is rarely a notion of "informed consent," i.e., explicitly informing the individuals about all the actions that will be undertaken using such data. Multiple individuals have expressed regret and raised concern upon realizing that their data has been used by machine learning algorithms for training tasks such as face detection.[1] Though the GDPR does not make processing dependent on consent, if consent is used as a processing ground, the GDPR requires that consent is both informed and freely given, and can be withdrawn as easily as given. Moreover, processing that is not necessary for a given purpose is unlawful, whether based on consent or one of the other legal basis. This implies that under the GDPR, repurposing of data processing may be illegal. This implies that an image posted on Facebook or elsewhere on the world wide web cannot be processed for a purpose that was not communicated to the data subject and for which no legal basis applies.

Control

As the AI elements in multimedia are entering myriad applications, an important question is whether an individual can actually decide and control how much AI is being used. For instance, should the users know that AI is used in a service, be it a natural language based chat bot or an image analysis software, which creates automatic captions for each photograph? Similarly, it would be important for users to have choices beyond the binary "install/don't install" and be able to control the degree to which automatic AI processing is part of the workflow. For instance, one user may want to allow AI in a video-conferencing application for facial identity analysis but not for face touch-ups and vice-versa. We need frameworks that support such a process workflow and make it easy for users to control their choices at different points of time. These are pivotal questions, directly related to human autonomy and dignity, especially in the light of attempts to "nudge" people into compliant behavior "behind their back"––for instance, based on emotion detection in facial images.[40]

Legal Compliance

With the growth of research impact outside the lab environment, there is a need for legal compliance at the level of the design of an application. Laws like the GDPR have made many of the above aspects a critical legal requirement rather than being "good things to do," notably by requiring a data protection impact assessment in case of likely high risk to fundamental rights and freedoms, and by requiring data protection

by design to mitigate such risks. These are legal obligations that level the playing field. For instance, imposing these duties on all companies that want to operate in the market, there can be strong economic incentives created for companies to pay keen attention to the consequences of deployment of AI. This should help research communities, such as in multimedia to come up with different types of research design that incorporate the consequences of design choices.

Problems Targeted

Multimedia research has paid significant attention to commercially viable applications such as ad placement and product recommendation but relatively much less effort to societally relevant but less directly marketable applications such as long-term support for education, healthcare or tackling climate change.[27] In fact, "social good" has been identified as a key focus area of multimedia research in a recent NSF workshop on Multimedia Challenges and a column by the Associate Editor-in-Chief in IEEE Multimedia urges researchers to be mindful of social impact of the applications being created.[27,42] Hence, it might be a good time for us as a community to introspect, and prioritize research on socially relevant themes in the coming decade.

EMERGING SOLUTION PATHS AND OPEN RESEARCH QUESTIONS

There are multiple approaches that are painting an optimistic picture regarding multimedia research in addressing each of the abovementioned issues. However, much more work is needed. Here, we identify multiple research questions and research areas that are ripe for exploration and development.

Use of Web-sourced Data for Large Dataset Creation in Relation to Data Protection

One of the biggest drivers for deep learning-based multimedia research is the recent availability of large-scale web-sourced image and video datasets. Clearly, not every image or video that is available on the web should be downloaded and used as part of a dataset. Notably, the GDPR always requires a valid legal ground and an explicit, specified and legitimate purpose for the processing of facial images as they concern identifiable data. Even consent is only valid when given for an explicit, specific, and legitimate purpose. The issue does not end there and there are ethical ramifications of using one's facial data, for say, profiling applications in the future. Although the GDPR has a broad research exception, this mostly applies to research in the public interest, rather than commercial interests. Multiple scholars have argued that even where the GDPR does not apply, just because the data is "public" does not mean it is acceptable for researchers or corporate agencies to reuse it for their purposes.[5] The default prohibition of automated targeting as codified in the GDPR squarely addresses this issue. Therefore an important question for the multimedia research community is to identify the guidelines for legal compliance and, within the space left open by the law, for ethically creating such large-scale datasets.

Multiple emerging efforts in multimedia research are now focusing on less data-hungry approaches for artificial intelligence. These approaches include the creation of domain-aware (e.g., physics inspired) approaches, zero and one-shot learning approaches, and transfer learning. (e.g.,[19, 20]) While domain (e.g., physics) inspired approaches clearly do not need lots of data to get started, other machine learning approaches are also trying to reduce the amount of new data needed to tackle each emerging problem. However, it is still early days in this space, and the legal and ethical requirements mentioned above clearly call for an important research direction––one on approaches that do not require large datasets.

Informed Consent and Control in Relation to Copyrighted Content and Portrait Rights.

Creative Commons provides an approach for identifying the permissions on what can be done with the images.[37] These licenses concern the copyright of the "author" of the image, not the person depicted in the image. However, Creative Commons licenses were defined before the deep learning and the corresponding opportunities for identifying individuals became commonplace, meaning that we can now assume that *insofar as facial images are concerned, those depicted have a so-called portrait right in the picture, which concern their privacy right rather than the photographers copyright.*

One of the possible settings considered during a panel discussion[43] on this topic at the 2019 ACM Multimedia conference was a "No AI" permission setting, which would make it illegal for algorithms to use the image for training of sophisticated face matching algorithms. However, the solution is not as simple as it appears. For instance, does the above "No AI" tag also include image cropping, touch-up, lighting, or other filters? When does the processing become "AI" is not one with a clear definition, and understanding the user's perception/understanding of what they are signing up for remains an important research problem. Understanding this would require work by those who not only understand the underlying technology but also understand the human perspective on these topics. Under the GDPR the issue plays out differently, as consent can only be provided for a specific purpose. Providing consent for "whatever" processing as long as it does not involve AI would be invalid.

Another point that generated large agreement in the panel discussion was providing users the ability to withdraw consent at any later point of time, as is now required under the GDPR. Note that under the GDPR, the mere fact that one has made public one's image does not imply consent for processing by whoever for whatever purpose. Some online systems have started designing web repositories (see OpenPDS, an open source personal data store[8]) that allow for users to remove their data at any point of time. However, in image and multimedia research the issue is more complicated. If there is a model that has learnt using millions of images, does the model also need to be discarded if the consent for (even one of) the supporting images is retracted? This question has informed the work on differential privacy,[38] which solves the problem to the extent that the model will not allow for reidentification. Under the GDPR, this would mean that the model does not qualify as personal data, and therefore the GDPR does not apply to the model. Again, there is a need for more research and identifying community norms in this space and the research findings could inform the legal viewpoint in this space.

Algorithmic Bias

Multiple studies have now accumulated evidence that computer vision and multimedia algorithms can be biased in terms of their performance across demographic groups. The reasons for these biases include the imbalance in training data sets, lack of positive training samples for historically marginalized communities, lack of training data to allow for convergence, and the lack of awareness regarding the leakage of demographic information (e.g., a "moustache detector" hidden in the layers of convolutional neural networks) in the developed algorithms. The default legal prohibition of indirect discrimination on grounds, such as gender and ethnicity may have unexpected repercussions when proxies are used that result in effective discrimination of women or ethnic minorities. Hence, an important question for the multimedia research community is how to develop multimedia algorithms that support both high accuracy and low bias?

Some of the possible approaches to counter this include those suggesting the use of datasets with balanced representation of people with different demographic characteristics––some of which may be artificially generated, creating adversarial approaches that penalize algorithms for any perceptible bias,[10] and those that propose posthoc adjustment of results for countering bias.[11] As an illustration of this kind of work in multimedia research, a recent paper by Alasadi *et al.*, describes a GAN (generative adversarial network) approach for face matching where one network optimizes for face matching, whereas another network tries to reduce bias. Specifically, the second network tries to infer demographic properties from the hidden layers of the first network and evidence of gender encoding (even when not directly required for the assigned task) is considered evidence of bias. The competition between the two networks yields models that balance accuracy and fairness.[10] We note, however, that since facial recognition systems are sometimes used for surveillance purposes that disadvantage specific groups, it may or may not be in the interest of those groups to become more identifiable. [39]

Explainability and Control of Algorithms

One of the side effects of the development of deep learning approaches is the complexity of the developed algorithms, which comes with the side effect of no human being able to explain the details of the algorithms developed in terms of the features being implemented or the decision rationale. This has costs in

terms of interpretability of the models and the lack of transparent causal reasoning for the decisions being made by the system. If such a system needs to make important decisions (e.g., in life and death scenarios in autonomous vehicles) then an explanation of the underlying processes is important. The GDPR requires that automated decisions that seriously affect people are accompanied by meaningful information about the "logic of processing," which implies that such decisions are prohibited if no meaningful information can be provided. Multiple research efforts in machine learning have started focusing on explainability in AI (See[12] for a review). One limitation is that current approaches might help experts to understand the inner workings of ML approaches, but they are of lesser usefulness to domain experts without any ML background. We need a user-centered perspective on the use of AI in applications and systems in which individuals can understand which information and decision is AI supported and if and how they can opt in or opt out.

Community Norms on Research

Given the wide variety (geographic, disciplines, political) of viewpoints represented within the ACM multimedia community, can there be a common set of guidelines that make sense to all researchers? *While inherently difficult, multiple disciplines ranging from nuclear physics to drug testing have come up with globally accepted guidelines for research.* Also, what should be the mechanism for supporting the development of such an ethical framework and how can such guidelines be implemented in the review process? Finally, how do we prevent discussing legal obligations as if they were ethical principles (often framed as "ethics washing"). There is a need for fundamental research as well as organized consensus-building within the multimedia research community to agree on a common set of norms that would be applicable across the globe. Some of these norms could be made part of the paper review/acceptance process in the community going forward. For instance, some communities require access to data to allow for replication of results before accepting research papers. Others insist that the paper cannot be accepted without a formal review by an ethics board. The multimedia research community has been pioneering some efforts on replication of results and perhaps the scope can be broadened to allow for dedicated benchmarks that allow authors and the wider community to reserve their research efforts for work that aligns with the most basic ethical norms (in the case that these norms are not already part of the applicable legal framework).

In summary, there is an urgent need to raise awareness about ethical and legal challenges in multimedia research. While there are multiple challenges, there are also opportunities to undertake meaningful research, which is technically robust and societally beneficial.

REFERENCES

1. R. Metz, If your image is online, it might be training facial-recognition AI, 2019. [Online]. Available: https://www.cnn.com/2019/04/19/tech/ai-facial-recognition/index.html
2. J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proc. Conf. Fairness, Accountability Transparency*, 2018, pp. 77–91.
3. Facial recognition poses serious risks. Congress should do something about it, 2018. [Online]. Available: https://www.washingtonpost.com/opinions/facial-recognition-poses-serious-risks-congress-should-do-something-about-it/2018/07/18/c4c8973c-89c1-11e8-a345-a1bf7847b375_story.html?utm_term = .8a991491dd14
4. San Francisco just banned facial-recognition technology, 2019. [Online]. Available: https://www.cnn.com/2019/05/14/tech/san-francisco-facial-recognition-ban/index.html
5. K. Albury, "Just because it's public doesn't mean it's any of your business: Adults' and children's sexual rights in digitally mediated spaces," *New Media Soc.*, vol. 19, no. 5, pp. 713–725, 2017.
6. Hamburg regulator bans Google from listening to smart speaker audio, 2019. [Online]. Available: https://globaldatareview.com/article/1195881/hamburg-regulator-bans-google-from-listening-to-smart-speaker-audio
7. Data for Good: FATES, Elaborated, 2018. [Online]. Available: https://datascience.columbia.edu/FATES-Elaborated
8. Y. A. De Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, "openPDS: Protecting the privacy of

metadata through SafeAnswers," *PloS One*, vol. 9, no. 7, 2014, Art. no. e98790.
9. M. Merler, N. Ratha, R. S. Feris, and J. R. Smith, "Diversity in faces," 2019, arXiv:1901.10436.
10. J. Alasadi, A. Al Hilli, and V. K. Singh, "Toward Fairness in Face Matching Algorithms," in *Proc. 1st Int. Workshop Fairness, Accountability, Transparency MultiMedia*, 2019, pp. 19–25.
11. P. K. Lohia, K. N. Ramamurthy, M. Bhide, D. Saha, K. R. Varshney, and R. Puri, "Bias mitigation post-processing for individual and group fairness," in *Proc. ICASSP IEEE Int . Conf . Acoustics, Speech Signal Process .*, 2019, pp. 2847–2851.
12. M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Commun. ACM*, vol. 63, no. 1, 68–77, 2019.
13. J. Angwin, J. Larson, S. Mattu, and L. Kirchner, May 23, 2016, "Machine bias," ProPublica.
14. S. Barocas and A. D. Selbst, "Big data's disparate impact," *Calif. L. Rev.*, vol. 104, 2016, Art. no. 671.
15. C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proc. 3rd Innov. Theoretical Comput. Sci. Conf.*, 2012, pp. 214–226.
16. A. Datta, M. C. Tschantz, and A. Datta, "Automated experiments on ad privacy settings," in *Proc. Privacy Enhancing Technol.*, vol. 2015, no. 1, 92–112, 2015.
17. N. Diakopoulos and M. Koliska, "Algorithmic transparency in the news media," *Digital Journalism*, vol. 5, no. 7, pp. 809–828, 2017.
18. "False Testimony," *Nature*, vol. 557, no. 7707, May 2018, Art. no. 612.
19. Y. Yang, Y. Luo, W. Chen, F. Shen, J. Shao, and H. T. Shen, "Zero-shot hashing via transferring supervised knowledge," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 1286–1295.
20. R. Stewart and S. Ermon, "Label-free supervision of neural networks with physics and domain knowledge," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 2576–2582.
21. W. Schreurs, M. Hildebrandt, E. Kindt, and M. Vanfleteren, Cogitas, ergo sum, "The role of data protection law and non-discrimination law in group profiling in the private sector." *Profiling the European. Citizen*. Berlin, Germany: Springer, 2008, pp. 241–270.
22. Ethics guidelines for trustworthy AI, 2019. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
23. Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk [Online]. Available: https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281
24. EPIC Promotes 'Algorithmic Transparency' for Political Ads. [Online]. Available: https://epic.org/2017/11/epic-promotes-algorithmic-tran-1.html
25. A new study finds a potential risk with self-driving cars: Failure to detect dark-skinned pedestrians, 2019. [Online]. Available: https://www.vox.com/future-perfect/2019/3/5/18251924/self-driving-car-racial-bias-study-autonomous-vehicle-dark-skin
26. R. Caplan, J. Donovan, L. Hanson, and J. Matthews, "Algorithmic accountability: A primer," *Data Soc.*, vol. 18, 2018.
27. S. F. Chang, et al., "Report of 2017 NSF Workshop on Multimedia Challenges, Opportunities and Research Roadmaps," 2019, arXiv:1908.02308.
28. L. D. Introna, "Disclosive ethics and information technology: Disclosing facial recognition systems," *Ethics Inform. Technol.*, vol. 7, no. 2, pp. 75–86, Jun. 2005.
29. M. Flanagan, D. Howe, and H. Nissenbaum, "Values in design: Theory and practice." *Information Technology and Moral Philosophy*, ed. Jeroen Van den Hoven and John Weckert, Cambridge, U.K.: Cambridge Univ. Press, 2007.
30. I. D. Raji, T. Gebru/ M. Mitchell, J. Buolamwini, J. Lee, and E. Denton, "Saving face: Investigating the ethical concerns of facial recognition auditing," Jan. 2020, *arXiv:2001.00964*.
31. M. Hildebrandt, *Law for Computer Scientists and Other Folk*. Oxford, U.K.: Oxford Univ. Press, 2020, [Online]. Available: https://lawforcomputerscientists.pubpub.org/
32. S. Barocas, M. Hardt, and A. Narayanan, fairness and machine learning, 2019, [Online]. Available: https://fairmlbook.org/
33. J. Powles, Medium: The seductive diversion of "Solving" Bias in artificial intelligence, Dec. 2018, [Online]. Available: https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53
34. European Commission, COM(2020)64, "Final report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, Feb. 2020 [Online]. Available: https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications

-ai-internet-things-and-robotics_en
35. M. E. Kaminski, "The right to explanation, explained," Rochester, NY, USA: Social Science Research Network, SSRN Scholarly Paper, Jun. 2018, [Online]. Available: https://papers.ssrn.com/abstract = 3196985
36. A. Cavoukian, "Privacy by Design and the Emerging Personal Data Ecosystem," 2012.
37. Creative Commons––About The Licenses. [Online]. Available: https://creativecommons.org/licenses/
38. C. Dwork, "Differential privacy," in *Automata, Languages and Programming* ed. M. Bugliesi, et al., vol. 4052, Berlin, Germany: Springer, 2006, pp. 1–12, [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id = 64346
39. S. Fussell, The strange politics of facial recognition," The Atlantic, Jun. 2019, [Online]. Available: https://www.theatlantic.com/technology/archive/2019/06/democrats-and-republicans-passing-soft-regulations/592558/
40. A. McStay, "Empathic media and advertising: industry, policy, legal and citizen perspectives (the Case for Intimacy)," *Big Data Soc.*, vol. 3, no. 2, Dec. 2016, Art. no. 2053951716666868.
41. J. Kobielus, "What does it mean to certify an ai product as safe?," 2018, [Online]. Available: https://www.dataversity.net/mean-certify-ai-product-safe/
42. A. Hanjalic, "Multimedia research: what is the right approach?," *IEEE MultiMedia*, vol. 24, no. 2, pp. 4–6, 2017, [Online]. Available: https://www.dataversity.net/mean-certify-ai-product-safe/
43. V. K. Singh, A. Hanjalic, E. André, S. Boll, M. Hildebrandt, D. A. Shamma, and T. -S. Chua, "Legal and ethical challenges in multimedia research," in *Proc. ACM International Conference on Multimedia*, 2019, pp. 2514–2515.

Evolving Career Opportunities Need Your Skills

Explore new options—upload your resume today

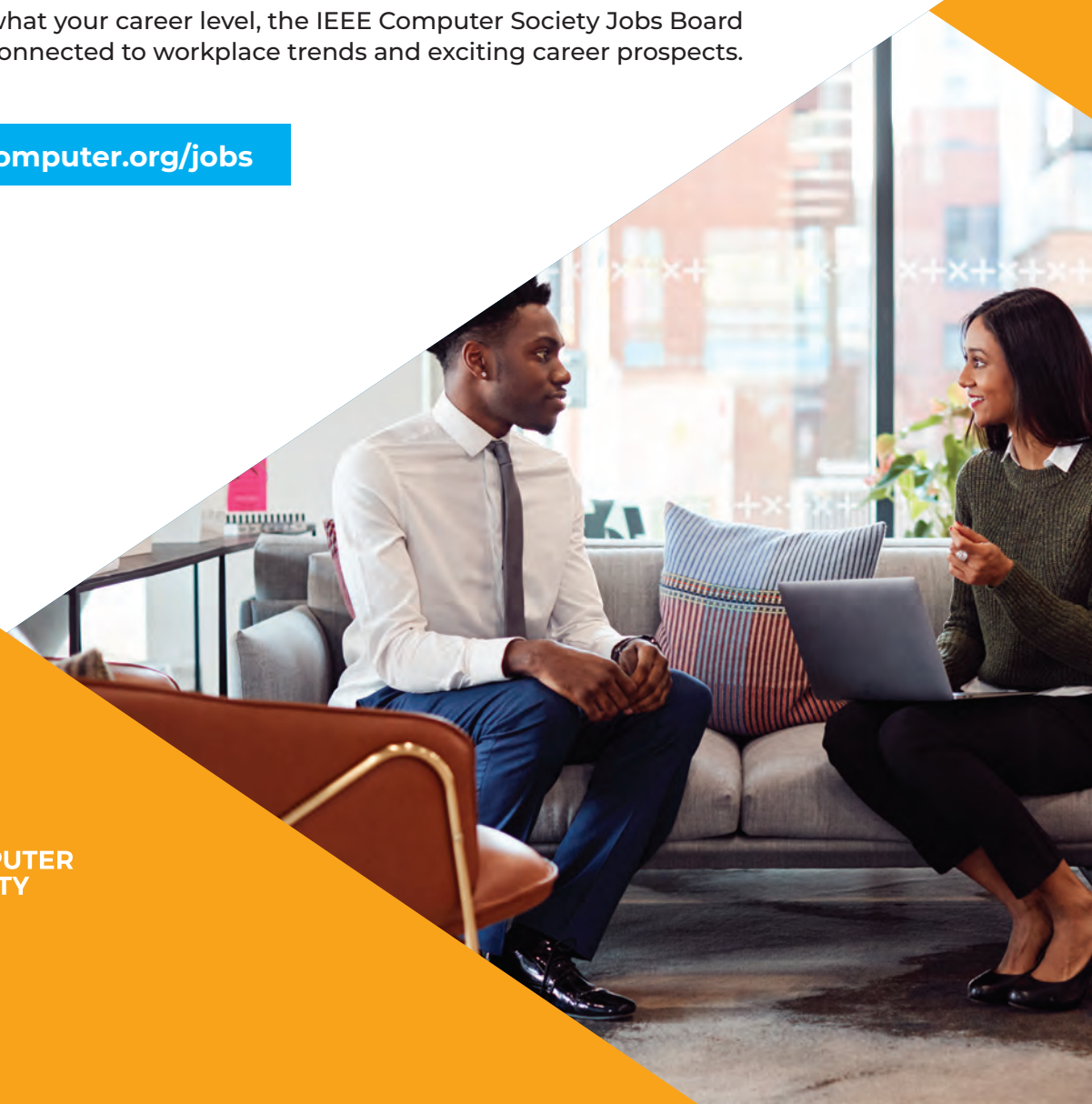
Changes in the marketplace shift demands for vital skills and talent. The **IEEE Computer Society Jobs Board** is a valuable resource tool to keep job seekers up to date on the dynamic career opportunities offered by employers.

Take advantage of these special resources for job seekers:

- JOB ALERTS
- TEMPLATES
- WEBINARS
- CAREER ADVICE
- RESUMES VIEWED BY TOP EMPLOYERS

No matter what your career level, the IEEE Computer Society Jobs Board keeps you connected to workplace trends and exciting career prospects.

www.computer.org/jobs



IEEE COMPUTER SOCIETY

IEEE

EDITORS: Michiel van Genuchten, VitalHealth Software, genuchten@ieee.org
Les Hatton, Oakwood Computing Associates, lesh@oakcomp.co.uk

This article originally appeared in
IEEE Software
vol. 37, no. 2, 2020

DEPARTMENT: IMPACT

Bringing Semantic Knowledge Graph Technology to Your Data

Bob van Luijt and Micha Verhagen

FROM THE EDITORS

Bringing the software to your data may, in times of privacy concerns, be a better approach than giving your data to the software companies. This is the intent of the knowledge graph described in this column. It is also the first column describing a product written in programming language GO. —*Michiel van Genuchten and Les Hatton*

The most used knowledge graph today is Google Search. What makes Google Search powerful is its ability to derive knowledge by finding a correlation between the concepts in a search query as opposed to a correlation between words (also known as *from strings to things*).[1] In a Google search, Google is able to increase the relevance of search results by making connections in its knowledge graph that exceed the possibilities of textual search.

As an example, ask yourself, "What concepts do you associate with the words *Steve Jobs*"—in this exact combination. You might think of Apple, Pixar, Steve Wozniak, or Buddhism. But someone else might have thought about a completely different person—the plumber down his or her street who is also called Steve Jobs. If you store information about both the entrepreneur Steve Jobs and the plumber Steve Jobs, the challenge is to map the right Steve to the context of the query.

The open source smart graph Weaviate (written in Go[2] and available through GitHub[3] and the Docker network[4]) aims to democratize similar semantic systems to business and community users so they can create their own scalable knowledge graphs. The biggest differentiator between Weaviate and solutions like Google Search and IBM Watson is that most of the data that organizations need for their crucial decision making are not publicly available or cannot be stored outside their own data centers for legislative, privacy, and ethical reasons. With Weaviate, organizations can set up their own knowledge graph, which means that sensitive and valuable data remain within the premises of the organization. Medical, financial, and government data are just a few examples that illustrate this point.

The enterprise version of Weaviate diverges from the open source version in that it comes with a different license, support, and service-level agreements, but the bits and bytes are the same. This structure allows small-time users like start-up entrepreneurs or academics to create high-end smart graphs similar to those created at large enterprises (Figure 1). Development started in 2016[5] and has had 12 committers, and almost 8 million lines of code have changed over time.[6] Table 1 shows that currently, the product consists of more than 90,000 lines of code. It is assumed that the overlap between the machine-learning model, data store, and software will have a bias toward the machine-learning model.

Digital Object Identifier 10.1109/MS.2019.2957526
Date of current version: 12 February 2020

 Published by the IEEE Computer Society

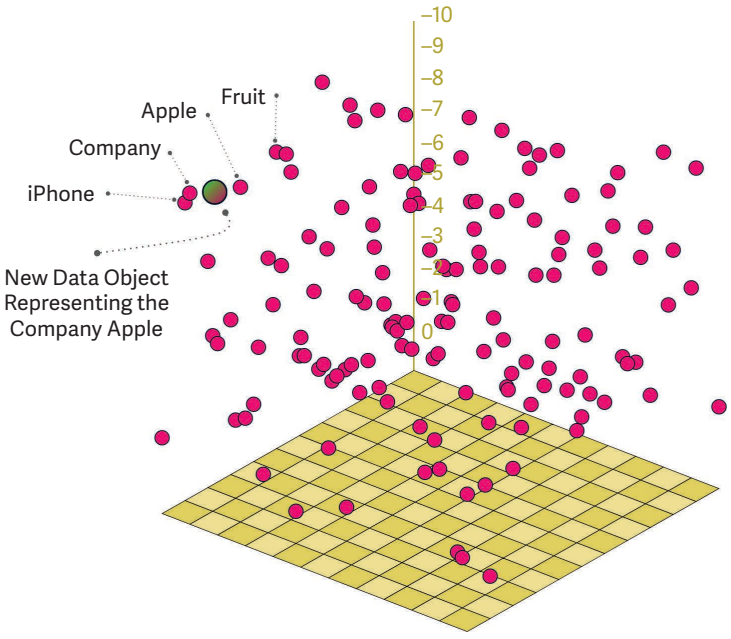


FIGURE 1. Contextionary visualization.

This column is comparable to some previous articles that described a product with an open source and a commercial license version (Bayesian networks,[7] WordPress,[8] and Eye[9]). The most comparable of the three is Eye, both in application domain (reasoning about qualitative information) and its installed base. The product may be relatively small in terms of size (lines of code); this can be attributed to the growing efficiency of the core programming language (Go) (Figure 2) and the continuous refactoring of the codebase (from its inception until today, almost 8 million lines of code have come and gone to reach the most optimized version).

In the 2016 article "Towards a Definition of Knowledge Graphs,"[10] the authors give a comprehensive definition of a knowledge graph: "A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge." When creating knowledge graphs with traditional graph databases, there is much focus on the graph (for example, through linked data and Resource Description Framework [RDF] structures), but our aim was to bring tools to developers that would help with the reasoning part by offering natural language processing through an intuitive RESTful and GraphQL[11] application programming interface (API) while maintaining speed and scalability.

TABLE 1. The size of the product.

Language	Files	Code
Go	698	71.725
JSON	37	15.124
JavaScript	5	3.112
YAML	14	795
Markdown	7	528
Bash	25	429
Groovy	2	28
Plain Tekst	1	12
Total	789	91.753

JSON: JavaScript Object Notation; YAML: YAML Ain't Markup Language; Bash: The Bourne-Again Shell.

THE CONTEXTIONARY

The core design principles of our smart graph are focused around a user experience-friendly API and scalability with one core focus in mind: enabling anybody to create his or her own high-end semantic knowledge graphs.[12]

```
func (cv *Vectorizer) occurencesToWeight(occs []uint64) []float32 {
        factor := cv.config.OccurenceWeightLinearFactor
        max, min := maxMin(occs)

        weights := make([]float32, len(occs), len(occs))
        for i, occ := range occs {
                // w = 1 - ( (O - Omin) / (Omax - Omin) * s )
                weights[i] = 1 - ((float32(occ) - float32(min)) / float32(max-min) * factor)
        }

        return weights
}
```

FIGURE 2. An example of calculating the occurrences in Go.[2]

```
{
  "Company": {
    "name" : "Apple Inc.",
    "foundedIn": 1976,
    "foundedBy": {
      "Person" : {
        "name": "Steve Jobs"
      }
    },
    "hasCeo" : {
      "Person" : {
        "name" : "Tim Cook"
      }
    }
  }
}
```

FIGURE 3. A JSON representation of Weaviate's class and property structure.

The first inspiration came from the article "From Machine Learning to Machine Reasoning,"[13] in which Leon Bottou argues (as explained by Yann LeCun on the Artificial Intelligence Podcast[14]) that a learning system should be able to manipulate (data) objects that are stored in a space and restore them when a new meaning can be attached to these data. Weaviate uses an enhanced and improved semantic vector–space storage model based on word-embedding techniques developed at Stanford University called *Global Vectors for Word Representation*.[15] A supermarket can be used as a metaphor to understand the structure and organization of the semantic vector space. If you go to a supermarket with a shopping list containing three items (an apple, a banana, and laundry detergent), when you locate the apple, you know that in the supermarket space, the distance from the apple to the banana will be less than the distance from the apple to the laundry detergent. This example is similar to how a vector space storage model based on word embedding works. Concepts that are correlated are stored close together. Weaviate doesn't store data in a traditional table structure or graph structure but in a spatial representation where data objects are placed together based on their semantic meaning.

This model is called the *contextionary*.[16] Every data object added to the knowledge graph is placed inside the vector space based on its semantic meaning. This means a user can query for data based on what the data are about (i.e., the context in which they are placed) rather than fixed keywords (Figure 3).

DEVELOPMENT

For use in an enterprise production environment, the contextionary has to be fast, scalable, and user friendly to be successful. Retraining of machine-learning models should not interfere with the user experience of the developer. Therefore, one of the key features is the ability to add a new data object to the contextionary without the need for retraining the model because knowledge graphs need to be fast. Retraining a model after the addition of new data objects would be unusable in real-life cases.

In the contextionary, this issue is solved by calculating new vector positions for data objects based on the semantic meaning of the group of concepts that

describe it. When one adds a new data object through the RESTful API—for example, a company with the name Apple—the contextionary generates a new vector position for this object algorithmically without the loss of semantic meaning. This is done by calculating a weighted centroid based on the known concepts and occurrences of the concepts in the original training corpus.

The vector position of a new object is determined in real time by taking the vector position of the data object and calculating the centroid based on the number of occurrences of the word in the original training corpus. The fewer times a word occurs, the more weight it is given.

The weight factor is added to prevent the centroid from gravitating toward the center of the contextionary. The center is something we want to avoid because, through experimentation, we learned that this is the place where ambiguity exists.

A second feature is the ability to deal with ambiguous words. Because the model works with individual words as tokens, a term with multiple meanings becomes ambiguous. For example, the token "apple" refers to a vector position related to both fruit and the iPhone. By moving the centroid of the search term toward other words (e.g., between "apple" and "company"), semantic placement of the data objects can be determined with varying accuracy. This same algorithm is used to search through the vector space while using multiple keywords.

Although the latest cutting-edge research in natural language processing (e.g., BERT, ELMo, and so on) uses bidirectional vector representations (i.e., a word is given a vector position based on the context rather than as a single representation of that term), Weaviate uses the unique representation of a word because it works as a map of meaning in which to place data objects.

The training of concepts is also incorporated. For example, the words "Steve" and "Jobs" often occur in relation to the company Apple. When adding data objects, we give commonly occurring combinations of words more weight when determining the vector position. This further increases the semantic accuracy of the data object. Figure 4 illustrates how these objects are represented; note that all known words have individual 600-dimensional vector representations, like "company," "name," "apple," "in," "founded," and so on.

```
{
    "synonym": "The Wonderful Wizard of Woz",
    "concept": "Nickname of Stephen Gary Wozniak,
                co-founder of Apple Inc."
}
```

FIGURE 4. A JSON example of adding new custom concepts. All known words have individual 600-dimensional vector representations, like "wonderful," "wizard," and so on.

A third important feature is the ability for organizations to add concepts to the contextionary that are specific to the environment in which they operate. The contextionary is trained on many open data sources such as Wikipedia, so it knows about common knowledge. However, a word combination that might not be known (e.g., "The Wonderful Wizard of Woz," a nickname for Steve Wozniak) is a concept on which the contextionary was not trained. To enable organizations to add domain-specific knowledge, we use the same methodology, similar to the way humans learn new concepts and nomenclature—by defining it in language that is already known. A vector position for the combination "The Wonderful Wizard of Woz" is calculated based on the words and concepts the knowledge graph already knows. Thus, when new data objects are added, the knowledge graph can store the new information based on the semantic meaning it now has for the word combination "The Wonderful Wizard of Woz."

FOCUS ON THE API USER EXPERIENCE

What most new technologies have in common is the ease of use. The two most important criteria for choosing a query language are adoptability (is there something we can leverage with a large community?) and simplicity (how can we avoid complex graph query languages?).

We considered SPARQL, the RDF query language developed and defined by the World Wide Web Consortium and available since 2008. Although SPARQL is loved in the academic community, the engineering community had a different opinion: that it is complicated and not very developer friendly. Therefore, we decided to adopt the new graph query language

```
{
  Get {
    Things {
      Company (
        explore : {
          concepts : [ "windows" ],
          moveAwayFrom : {
            concepts : [ "technology" ],
            force : 0.9
          },
          moveTo : {
            concepts : [ "glass" ],
            force : 0.85
          }
        }
      ) {
        name
      }
    }
  }
}
```

FIGURE 5. Weaviate's GraphQL representation.

GraphQL, an open source data query and manipulation language that was started at Facebook but is now part of the Linux Foundation. The use of GraphQL comes with a tradeoff; it is less expressive but much easier to use and adopt. There are dozens of libraries available in many languages, and it is easy exposable through the RESTful API (Figure 5).

Exploration syntax is based on GraphQL, which makes it extremely simple to request data and explore the graph. An example of a query is based on a demo data set that is supposed to return names of companies listed on the New York Stock Exchange. Exploring only "windows" would most likely return "Microsoft Corporation," but by moving away from the concept of "technology" and toward "glass," it is more likely that producers of glass windows will be given.

USE CASE: AUTOMATIC CLASSIFICATION AND SEMANTIC SEARCH THROUGH CYBERSECURITY THREATS

Knowledge graphs can be employed for many use cases where semantic search, automatic classification, or knowledge representation plays a role. A good example to illustrate all three functionalities is a cybersecurity use case based on the Mitre ATT&CK Framework,[17] a vast array of different types of cybersecurity threats with corresponding mitigation recommendations. However, mapping messages (e.g., classifying that an email containing information about somebody asking for a password is a phishing attack) is a task that is often still done manually or based on fixed keywords. Based on a description of observations and system inputs, Weaviate enables the automatic identification and classification of the most likely type of attack and recommends corresponding countermeasures.

The semantic search functionality allows searching through vast amounts of security tickets for context rather than keywords alone. For example, phishing-type tickets can be found by querying for "fraudulent attempt to obtain sensitive information" without that relation ever being explicitly made. The words in the query can be found in a position nearby the word "phishing."

Cybersecurity is just one of the many domains where knowledge graphs can be applied. Semantic search, automatic classification, and knowledge representation are new frontiers within data management because the problems they solve are currently handled manually by creating complex taxonomies or labeling data sets.

WHAT THE FUTURE MAY BRING: KNOWLEDGE NETWORKS

The next step is the development of a knowledge network peer-to-peer protocol enabling the semantic exploration of concepts across organizations. The protocol we aim to develop determines how peers can securely explore and combine information from their own knowledge network with information stored in Weaviate instances that belong to other peers. A peer can be any entity, such as a university, business, or individual.

Because a query can be agnostic of the language used, a network of Weaviate instances, hosted over the World Wide Web, can be queried without the questioner knowing what keywords to use in advance. Networks can be public or private and hosted within an organization's own digital walled garden or over multiple datacenters.

We see many opportunities to use the power of knowledge graphs without having to surrender your knowledge to software giants that may exploit that knowledge in unknown ways.

REFERENCES

1. Google, "*Introducing the knowledge graph: Things, not strings*," May 16, 2012. [Online]. Available: https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html
2. Wikipedia, "*Go (programming language)*." Accessed on: July 17, 2019. [Online.] Available: https://en.wikipedia.org/wiki/Go_(programming_language)
3. SeMI Technologies, "*Weaviate: Open source knowledge graph (GraphQL/RESTful/P2P)*," GitHub. Accessed on: May 5, 2019. [Online]. Available: https://github.com/semi-technologies/weaviate
4. SeMI Technologies, Amsterdam, North Holland, "*Installation guide*." Accessed on: Aug. 17, 2019. [Online]. Available: https://www.semi.technology/documentation/weaviate/current/installation.html
5. Weaviate repository, GitHub. Accessed on: Aug. 17, 2019. [Online]. Available: https://api.github.com/repos/semi-technologies/weaviate
6. Weaviate, "*November 29, 2019–December 6, 2019*," GitHub. Accessed on: Sept. 12, 2019. [Online]. Available: https://github.com/semi-technologies/weaviate/pulse
7. N. E. Fenton and M. Neil, "Decision support software for probabilistic risk assessment using bayesian networks," *IEEE Softw.*, vol. 31, no. 2, pp. 21–26, 2014. doi: 10.1109/MS.2014.32.
8. J. Cabot, "WordPress: A content management system to democratize publishing," *IEEE Softw.*, vol. 35, no. 3, pp. 89–92, 2018. doi: 10.1109/MS.2018.2141016.
9. R. Verborgh and J. De Roo, "Drawing conclusions from linked data on the web: The EYE reasoner," *IEEE Softw.*, vol. 32, no. 3, pp. 13–17, 2015. doi: 10.1109/MS.2015.63.
10. L. Ehrlinger and W. Wöß, "Towards a definition of knowledge graphs," in *Proc. SEMANTICS 2016*, Sept. 13–14, 2016, Leipzig, Germany. [Online]. Available: http://ceur-ws.org/Vol-1695/paper4.pdf
11. GraphQL, "*A query language for your API*." Accessed on: Oct. 21, 2019. [Online]. Available: https://graphql.org/
12. SeMI Technologies, Amsterdam, North Holland, "*Introduction to Weaviate*." Accessed on: Oct. 31, 2019. [Online]. Available: https://www.semi.technology/documentation/weaviate/current/
13. L. Bottou, "From machine learning to machine reasoning." *Machine Learning*, vol. 94, no. 2, pp 133–149, Feb. 2014. doi: 10.1007/s10994-013-5335-x.
14. Y. LeCun, "*AI podcast*," Aug. 31, 2019. [Online]. Available: https://www.youtube.com/watch?v=SGSOCuByo24&feature=youtu.be&t=1212
15. J. Pennington, R. Socher, and C. D. Manning, "*GloVe: Global Vectors for Word Representation*." Accessed on: Nov. 5, 2019. [Online]. Available: https://nlp.stanford.edu/projects/glove/
16. SeMI Technologies, "*Contextionary*," GitHub. Accessed on: Aug. 17, 2019. [Online]. Available: https://github.com/semi-technologies/contextionary/blob/081cc03160b65c59eddd185df94fa0c4f6cd3224/server/corpus_vectorizer.go#L137-L148
17. Mitre, McLean, VA, "*ATT&CK*." Accessed on: May 12, 2019. [Online]. Available: https://attack.mitre.org/

BOB VAN LUIJT is the product, commercial, community lead, cofounder, and first committer to Weaviate at SeMI Technologies, Amsterdam, The Netherlands. Contact him at bob@semi.technology.

MICHA VERHAGEN is the finance and operations lead and cofounder at SeMI Technologies, Amsterdam, The Netherlands. Contact him at micha@semi.technology.

This article originally appeared in
IEEE micro
vol. 40, no. 2, 2020

Expertise at Our Fingertips

Shane Greenstein, *Harvard Business School*

When my brother-in-law moved out of state, he gifted my household his beautiful leather-bound set of Encyclopedia Britannica. In spite of their age, they contain answers to any number of questions. How many species of penguins inhabit Antarctica? Who was husband to Cleopatra, last queen of Egypt? When was Billie Holiday born?

Nobody in my household ever touches them. Everybody uses Wikipedia.

THIS COLUMN CONTRASTS THE ECONOMICS BEHIND YESTERDAY'S COMPENDIUM OF EXPERTISE AND TODAY'S CROWD-SOURCED WIKI. ANY COMPARISON, EVEN A COARSE ONE, WILL SHOW THAT PRICES FELL DRAMATICALLY.

This column contrasts the economics behind yesterday's compendium of expertise and today's crowd-sourced wiki. Any comparison, even a coarse one, will show that prices fell dramatically. No other conclusion can emerge. Did quantity and quality of answers improve? How about their accuracy and reliability? That is less obvious.

CHEAPER PRICES

Start with prices. Britannica reached its peak sales in 1990 when a set of books cost a household around $1500, just under $3000 in contemporary 2020 dollars. The leather-bound volumes cost 30% more. Most households purchased these with monthly payments, say, $30–$50 a month. That was 1%–2% of median U.S. family income. Rich and well-to-do middle class families bought them.

The price of Wikipedia differs substantially. The ungated web site costs nothing to use, though it is also not entirely free to users. Think of it this way. Users pay charges for internet access. A portion of that expenditure anticipates using Wikipedia.

It cannot add up to much. For most households Wikipedia constitutes much less than 4% of their surfing. The average broadband subscription and smartphone data contract in the United States is around $40–$60 a month and around $60–$80 a month, respectively. The portion of expenditure attributable to Wikipedia cannot exceed $3–$5 a month. Anyway you look at it, expenditure declined more than 90%, and became a trifling fraction of median household income.

That does not account for the biggest drop, which does not involve money. It concerns time.

In 1990, Britannica sold more than 100,000 copies in the United States. Over several decades several million households bought and owned a volume of books. Tens of million households had a set from a competitor—e.g., World Book, Colliers, and so on. Everybody else went to their local library. In contrast, today, four out of five U.S. households access the internet at home, or on their phone, which means anywhere. If a price could be put on convenience, it would show a massive drop because internet access is so widespread.

Because Wikipedia uses less time, it is just less hassle per transaction. That enables a crazy difference in the scale of use. In one hour Wikipedia receives 4.8 million visitors for content in English. Not all of those readers came from the United States, but so what? One hour.

Scale changed in other ways. Britannica contains

Digital Object Identifier 10.1109/MM.2020.2971915
Date of current version 18 March 2020.

 Published by the IEEE Computer Society

120 000 articles at most. To facilitate sales of books, Britannica's managers decided long ago to cap the total volume of space occupied by the books on shelves. Editors shortened the included articles to make room for the new. That has not changed much over the decades.

Wikipedia contains a broader scope, and more information. *Due to* the negligible cost of storage Wikipedia faces no constraint on its breadth or length. At last count Wikipedia contains just under six million articles in English, and it continues to grow. For example, the entry for Penguins—i.e., the animal—contains more than eight thousand words, while the entry for The Penguin—i.e., the villain from Batman—gets five thousand. Cleopatra's entry receives more than 25 000 words. Billie Holiday gets more than 10 000, while Billie Eilish, recent Grammy winner at age 18, has 5000 words on her page.

Britannica does not lose on every dimension. Constraints led Britannica to impose a singular sensibility on all articles. Everything came from an established expert. Every article got attention from professional editors, so the best writing is truly magnificent. Wikipedia goes to the other extreme. While broad, it applies a porous filter. It contains a massive sampling of material that Britannica does not. It has something for everyone, and plenty that many readers do not want. After all, one expert's junk is another reader's appropriate topic for an online encyclopedia.

That is a crucial difference, so let's reiterate. Britannica has great content, and most of it is inconvenient to access. Wikipedia contains plenty of everything, both nutritious news and saccharine sweet sophistry, all of it within reaches with little effort.

SUPPLY OF AUTHORITY

The costs of production declined too. Britannica's expenses in 1990—i.e., to support worldwide distribution—reached $650 million, around $1.3 billion in today's dollars. Wikipedia's expenses—i.e., again, to support worldwide distribution—reached less than $100 million. Wikipedia costs at least 90% less to produce.

By design, and out of necessity, Britannica was selective. It showed only the final draft of an article. Also by design and out of necessity, Wikipedia is a work in progress. It shows everything. The cost of an extra web page is negligible, and so is the cost of another paragraph, picture, link, and an article's entire editorial evolution. The only effective constraint comes from the AI bots that rid the website of abusive language, and from the editors' collective sense of what belongs.

How does Wikipedia save so much production cost? Those editors are volunteers. The total of 250 000 of them regularly edit the site each month in all its languages. They add content, and incorporate suggestions from tens of millions who add something small. More than 350 employees support them.

Can volunteers write as well as experts? As it turns out, sometimes yes, when the crowd is big enough.

BRITANNICA HAS GREAT CONTENT, AND MOST OF IT IS INCONVENIENT TO ACCESS. WIKIPEDIA CONTAINS PLENTY OF EVERYTHING, BOTH NUTRITIOUS NEWS AND SACCHARINE SWEET SOPHISTRY, ALL OF IT WITHIN REACHES WITH LITTLE EFFORT.

Two colleagues and I recently investigated editors' behavior in the most uncomfortable setting at Wikipedia, the pages for U.S. politics.[1] We found that many editors start off biased. They show up and spout their slanted opinion. Unlike the majority, a future editor sticks around, at least for a short while. Of those, who remains for the long haul? We estimate that only 10%–20% stay for more than a year. Most leave after a month or so, most often after encountering others with extreme and opposing views. Most interesting, those who stay lose their biases, and start fostering a neutral point of view, the site's highest aspiration for all its content. In short, Wikipedia does not devolve into hopeless arguments because the moderates decide to stay and edit the crowd.

In another project, we compared the political slants of close to four thousand articles from Britannica and Wikipedia.[2] The articles covered nearly identical topics in U.S. politics, and tend to be popular. The biases of the articles in Britannica and Wikipedia became similar when the Wikipedia article received considerable editorial attention. The most edited achieved something akin to a neutral point of view. They were

almost always longer too, containing a wider sampling of opinion.

Articles matched by topic are not representative of all of Wikipedia, however. Due to its broad sampling of topics, Wikipedia contains many more articles, unmatched to anything within Britannica. But it comes with a drawback. Many of these lack editing, which generated the potential for uncorrected grammatical error, factual mistakes, and narrow sampling of opinion.

Therein lies the subtle difference between expert and crowd. The distribution of unfinished articles is enormous at Wikipedia because, believe it or not, 250 000 editors is nowhere near what the site requires. There are plenty of passages that need attention and do not get it.

WHILE WIKIPEDIA EXPOSES THE ARTIFICIAL ILLUSION OF USING A SINGLE SOURCE OF EXPERTISE, IT LEAVES THE CROWD'S AUTHORITY OPEN TO SECOND GUESSING. THE READER GAINS CONTROL, BUT LOSES ASSURANCE IN THE EXCHANGE.

Each organization acts accordingly. Britannica flaunts its expertise, while Wikipedia flaunts its sourcing from the crowd. Britannica claims reliability, while Wikipedia openly declares *caveat emptor*, recommending that anybody double check the answer against other sources.

Do readers check? Why should they check objective, verifiable, and noncontroversial facts? For example, there are eight species of penguins in Antarctica, Cleopatra's last husband was Marc Antony, and Billie Holiday was born on 1915. It took one editor little time to enter that information. If the first draft erred, somebody fixed it long ago. At Wikipedia the metadata for dates, numerical descriptions, historical accounts, and minutia of science shout their own accuracy.

On the other hand, neither Britannica, nor Wikipedia, can escape subjective, nonverifiable, and controversial content. How good was Burgess Meredith in his campy performance as the Penguin? Was Cleopatra considered charismatic? Why does Billie Eilish's music and fashion appeal to young listeners? Whereas Britannica retained its authority by asking readers to defer to the expert's opinion, Wikipedia invites answers for such questions from many sources, and tells readers to check elsewhere too.

While Wikipedia exposes the artificial illusion of using a single source of expertise, it leaves the crowd's authority open to second guessing. The reader gains control, but loses assurance in the exchange.

CONCLUSION

When questions come up in my household, I go straight to Wikipedia. It takes more time to put on reading glasses than it does to voice the answer from the small screen. The children listen while pretending not to, and so we move forward.

What a bountiful, convenient, and dangerous gift for the generation tied to small screens.

The user is in charge, but it comes with a catch. A modern reader needs to take the time to don their thinking cap. But who takes the time and effort? And does anyone really possess enough judgment to second guess it all?

REFERENCES

1. S. Greenstein, G. Y. Gu, and F. Zhu, Forthcoming, "Ideological segregation among online collaborators: Evidence from Wikipedians," *Manage. Sci.* [Online]. Available: http://dx.doi.org/10.2139/ssrn.2851934
2. S. Greenstein and F. Zhu, "Do experts or crowd-based models produce more bias? Evidence from encyclopedia Britannica and Wikipedia," *MIS Quart.*, vol. 42, no. 3, pp. 945–959, 2018.

SHANE GREENSTEIN is a professor at the Harvard Business School. Contact him at: sgreenstein@hbs.edu.

Get Published in the New *IEEE Open Journal of the Computer Society*

Submit a paper today to the premier new open access journal in computing and information technology.

Your research will benefit from the IEEE marketing launch and 5 million unique monthly users of the IEEE *Xplore*® Digital Library. Plus, this journal is fully open and compliant with funder mandates, including Plan S.

Submit your paper today!
Visit www.computer.org/oj to learn more.

IEEE COMPUTER SOCIETY | IEEE

This article originally appeared in
IEEE Internet Computing
vol. 22, no. 6, 2018

Teaching Crowdsourcing: An Experience Report

Hui Guo, Nirav Ajmeri, and Munindar P. Singh, *North Carolina State University*

Crowdsourcing is the process of accomplishing a task by using a typically open call to invite members of the public (the "crowd") to work on one's task. The authors describe a project assignment in which students received the opportunity of practicing crowdsourcing to accomplish a hummed song recognition task, yielding improved comprehension of the concept and high student satisfaction.

Crowdsourcing is about marshaling human knowledge and intelligence to solve tasks that are natural for humans but cannot be effectively performed by a computer. In addition to the idea of using humans, a distinguishing feature of crowdsourcing is to distribute the work to multiple humans, potentially including those with no specific credentials other than common sense, and usually from outside of one's organization. The term "crowd" is applied to the human workers to indicate that they may be nonspecialists and are selected from the public, although in practice the workers are chosen from eligible pools and may possess special knowledge or credentials. Leading applications of crowdsourcing include data gathering and analysis, crowdfunding, and idea generation. Although the idea of outsourcing work to the public dates back centuries, it was not until recently that distributing computations through microtasks became important in computing practice.[1] This paper is about the latter, narrower sense of crowdsourcing as a computing paradigm.

This paper describes an approach to teaching crowdsourcing with a practical orientation via an assignment we incorporated in our social computing course, which is offered to a mixture of graduate and undergraduate computer science students. Our course provides an introduction to the rich variety of social computing applications and identifies the concepts for their modeling and realization. Crowdsourcing is a key topic in social computing. Our instructional objectives were to familiarize students with the basic principles and logistics of crowdsourcing projects and introduce important elements of human computation, e.g., ways to motivate participation, such as deploying dynamic, adaptive, and personalized rewards (including incentives in terms of money and recognition).[2] The assignment we adopted was aimed to provide students hands-on experience on the technology, with respect to designing, deploying, and analyzing their own crowdsourcing projects and responses, as well as participating as a crowd worker for other projects. Deeper aspects of crowdsourcing, such as the differences between quality evaluation strategies, incentive mechanisms, and task setup, were introduced, but not within the scope of our learning goals. Our survey of the students showed a significant increase in their understanding of the concepts and workflow of crowdsourcing projects after our lectures and assignment.

Crowdsourcing has been taught in colleges. Instructors usually introduce crowdsourcing by adopting the requester role and using the students as a crowd. Davidson[3] introduced crowdsourcing to students by outsourcing part of the grading and teaching to students and received positive results. Proper peer grading, which we incorporate into our approach,

 Published by the IEEE Computer Society

has been proven to be beneficial to instructors and students.[4] However, we think it is imperative that the students practice this technique as requesters via commercial platforms, such as Amazon Mechanical Turk (MTurk), so that they are better prepared for when they need to employ crowdsourcing techniques. One of our students reported that his internship involved crowdsourcing and a proper exercise in class would be valuable for preparing for his work. Bigham *et al.*[5] argue for the importance of students participating in crowdsourcing projects both as requesters and crowd workers. They organize multiple projects throughout their crowd programming course where students experience aspects of crowdsourcing separately. Compared to their course, the time and budget dedicated to crowdsourcing in ours are much more limited.

Organizing a project assignment on crowdsourcing can be challenging. Our previous course assignments were primitively designed in that students were not inherently encouraged to take the task seriously and there was not a natural way for managers to evaluate the quality of work. Therefore, result aggregation phases did not necessarily achieve high quality and student did not report high satisfaction. In addition, we did not adopt a popular crowdsourcing platform. The challenges lie in the choice of suitable tasks and logistics during the process of the assignment. Students need to accomplish a computational task by using a crowd, which means the task should not be easily solvable by automated tools or students' own expertise. Also, crowdsourcing should be able to produce reliable and satisfying results for all or most students, which imposes requirements on both the quality of the specific task and the targeted crowd, especially with no funds being allocated. We propose that students use their classmates as the crowd, which requires the instructors to monitor the progress of the assignment closely, and resolve any logistics problems promptly. It can be nontrivial to make sure all students have a smooth and enjoyable experience when they have to act as employers, workers, and analysts. As part of our assignment, students exercised the workflow of a typical crowdsourcing project on a real-world platform. Doing so involved setting up a task, obtaining information from the crowd, rating and rewarding workers, analyzing responses, and computing an answer. They met the challenges arising in each of the above components. Most students reported that this assignment was interesting and educational, as well as valuable to their learning of crowdsourcing.

TASK AND SETTING

Amazon MTurk is a popular crowdsourcing platform, widely used for data collection or as a subject recruitment tool for behavioral[6] and political sciences.[7] MTurk is, therefore, an appropriate platform for instruction. We adopted MTurk Sandbox, a closed variant of MTurk that provides the same interfaces for managing expenses, rating answers, giving rewards but without payments—and hence is easier logistically. We instructed students to serve both as workers for MTurk human intelligence tasks (HITs) and as requesters using others as workers.

In order for students to experience the power of crowdsourcing, each student's assigned task should be relatively easy for a crowd of students but hard for an individual one. Also, each microtask should be enjoyable, since each student may be tasked to finish a large number of microtasks. The task we adopted in this assignment—namely, the identification of music from snippets—is well suited to crowdsourcing.[8] More specifically, we chose hummed song recognition as the target task.

Hummed song recognition challenges current computational techniques. First, the absence of lyrics makes it difficult to search song names. Wang *et al.*'s[9] query by a singing/humming system can misclassify

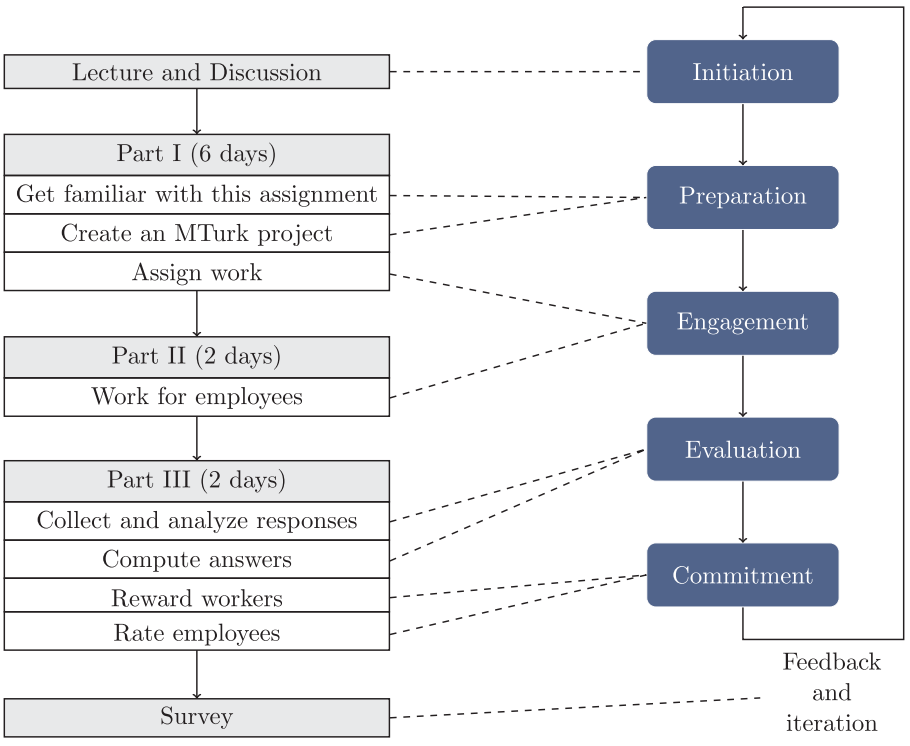


FIGURE 1. Crowdsourcing lifecycle[14] and assignment parts.

humming clips as singing, leading to erroneous output. Second, the accuracy of song recognition depends on the comprehensiveness of the existing dataset. Music recognition algorithms, e.g., Shazam,[10] identify snippets of existing recordings. Item-to-item similarity calculations that accommodate differences in timing and tempo scale poorly to datasets of millions of instances.[11] Third, since hummed recordings vary across users, extracting audio features that accurately represent music content remains challenging.[12] For example, Yang *et al.*[13] use absolute pitch values, which produce inaccurate results when people use dissimilar start tones. These challenges make crowdsourcing an appropriate approach for hummed song recognition. In our setting, hummed song recognition through crowdsourcing produced high accuracy for most students.

STRUCTURE OF THE CROWDSOURCING ASSIGNMENT

To lower the probability of one person being able to correctly recognize all songs, we generated a list of 1000 songs, of which 200 are classical music pieces, such as Beethoven's *Für Elise*; 200 are old favorites (for holidays or for children, such as *Santa Claus Is Comin' To Town*); and 600 are pop music. We sampled 200 from the 1000 songs, covering these classes with the same ratios. We recorded a humming of 10–20 s for each song, with the intention of producing recordings that would be easy to recognize. These recordings involved recognizable parts of songs such as the opening or chorus. A worker should be able to determine whether he or she recognizes the song instantly.

We published the 1000 song names, each with a numeric ID, so that only IDs could be used subsequently, which avoids the challenges of textual input, such as variability in input. Each student was assigned links to five recordings, with the ID of one of them revealed, and was asked to identify the remaining four. Students would crowdsource tasks to their fellow classmates.

Students act as both requesters and workers. One individual student's success requires the student's devotion as well as the timely cooperation from his or her classmates. We broke the assignment into three parts with different deadlines to make sure that all

students kept up with the schedule. In Figure 1, the three parts followed Kamoun *et al.*'s[14] five-stage life-cycle for crowdsourcing, namely, initiation, preparation, engagement, evaluation, and commitment. We set three deadlines to motivate students' participation throughout the workflow and to discourage them from putting off work until the end.

As part of the initiation phase, before this assignment, students had learned in class about the merits, basic workflow, and difficulties of crowdsourcing. We had given four lectures on crowdsourcing and other topics related to human computation, in which we had offered students extensive information on the theoretical background of gathering information from and outsourcing computation to a crowd of people. For example, we discussed in length the merits of negative surveys against positive surveys, the concept and examples of vox populi, a social mobilization example, gamification of certain human computation tasks, and so on. Students were given a list of relevant papers and notes as reading materials. Additional information regarding our teaching can be found here: go.ncsu.edu/teaching-crowdsourcing. Upon the announcement of this assignment, we explained the rationale of using crowdsourcing for hummed song recognition. To enable coordination, we provided a Google Sheet available to all students.

Part I: Create a Project

Students became familiar with the assignment and understood activities and deliverables. Each student was required to complete the following in six days:

- Set up a survey using a Google Form for workers, authenticated at the university, to give answers (IDs) for the hummed snippets.
- Set up a project on MTurk Sandbox, linking to the aforementioned Google Form and advertise it on the Google Sheet with a link to his or her MTurk HIT to recruit workers. For uniformity, we limited each project to a budget of $10 in notional money.
- Sign up as a worker for other projects on the Google Sheet. Select at least ten projects, taking projects that currently have fewer than ten workers (as a way to ensure all projects receive some workers).

We encouraged the students to give thoughts to the qualification strategies. Students were free to design their own tasks; they could ask the questions in any way and include any additional questions. We also encouraged them to advertise their projects on the course's online message board, which was public to the whole class.

Part II: Work for Others

Since each student was required to work for at least 10 projects and each project contained five recordings, each student worked on at least 50 recordings. A student submitted responses, i.e., recognized song ID for each recording, on the Google Form, and completed the HIT on MTurk Sandbox for each project.

Part III: Solution and Closing

Students produced song IDs for four recordings based on the answers they received on their projects within two days.

- Close the MTurk project.
- Gather responses on their Google Forms.
- Evaluate the responses, optionally using the song with the known ID for qualifying answers. Give due rewards to workers, summing up to $10.
- Analyze the responses to determine song IDs for the four recordings.

Project reports asked for

- Statistics and quality of answers received;
- Report on the analysis, including a) the method of computing IDs, and b) computed IDs of four song recordings;
- Rewards given to each worker, summing up to $10;
- Ratings of their employers, summing up to 10, or 1000%; and
- Comments and suggestions.

We encouraged students to write their own software for analysis and include their source code in their submission.

We asked the students to give ratings of their employers, as a way to compute each employer's

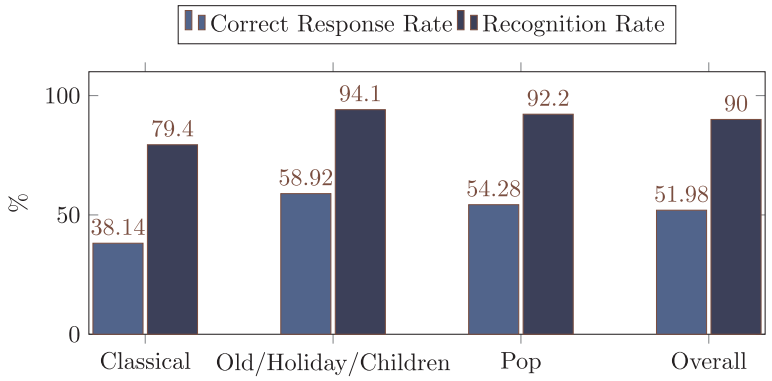


FIGURE 2. Accuracy of hummed song recognition.

reputation, similarly to how they would give rewards. The average reward for a worker from one employer is $1, and the average rating for an employer from one worker is 1 or 100%. Students were graded based on their performance in each part, incorporating the reward they received as workers and ratings as employers. We encouraged them to complete a post-survey regarding this assignment after they had completed it.

RESULTS AND DISCUSSION

A total of 34 students participated in this assignment, crowdsourcing the recognition of 170 recordings. Most of them were graduate students majoring in computer science, and approximately 20% were undergraduate students. On average, each student worked for 12 others, meaning that each recording received 12 answers. A total of 30 participants recruited 10 or more workers. A total of 27 students completed a post-survey (not everyone answered every question) about their experience.

We now present recognition results produced by our students, their performance and experience with the assignment, and recommendations for improvement.

Hummed Song Recognition

The accuracy of song recognition resulted from a combination of the quality of our recordings and the students' ability to identify them. Of the 170 recordings, 163 (96%) received at least one correct response. If we use the mode response for a recording, 153 (90.0%) were correctly recognized, which confirms that the recordings were of high quality. The actual recognition rate was slightly higher, since some students used their own expertise or additional validation (e.g., listening to all proposed songs to break a tie, which was a violation of crowdsourcing. We discuss this aspect of the findings below).

Of all the responses received from workers, only 52% were correct. Not surprisingly, incorrect responses were generally independent; hence, the correct response was likely to be the mode, if not the majority. A total of 16 recordings were correctly recognized with only two correct responses. Figure 2 shows the results for different classes. Overall, students accomplished the task with high recognition rates, which justifies the adoption of crowdsourcing for this task.

Notice that this success may not translate to real life. A real-life recognition task may include hard-to-identify, poor quality snippets. A list of possible answers, albeit a long one, that was provided to the workers could have influenced the results.

Assignment Execution

A few students failed to meet the deadline of Part I. It is not uncommon for students not to read the instructions until a few days before the final deadline. This assignment was of an unusual structure, with three separate deadlines. Some students did not realize the first deadline was actually four days before the final one. We emailed these students and they managed to keep up after that.

Most students did not make the effort of recruiting workers, although we had emphasized that recruiting was an important phase of crowdsourcing.

Since we were using their classmates as the targeted crowd, students were unaware of the possibility of lacking workers. In fact, the initial assignment of workers at the beginning of Part II was highly imbalanced, which we discuss later.

The average grade of this assignment was 94.5%. Most students conducted good analyses regardless of the quality of the responses. Students' grades included the rewards and ratings they received from their employers and workers, respectively. Nearly all students reported that the rewards and ratings were fair. The average reward and reputation each student received were $10 and 10, respectively, per the instruction, but the standard deviation of the rewards was $5.5, much higher than that of the ratings, which was 2.6. Students' effort as workers was much more varied than as requesters.

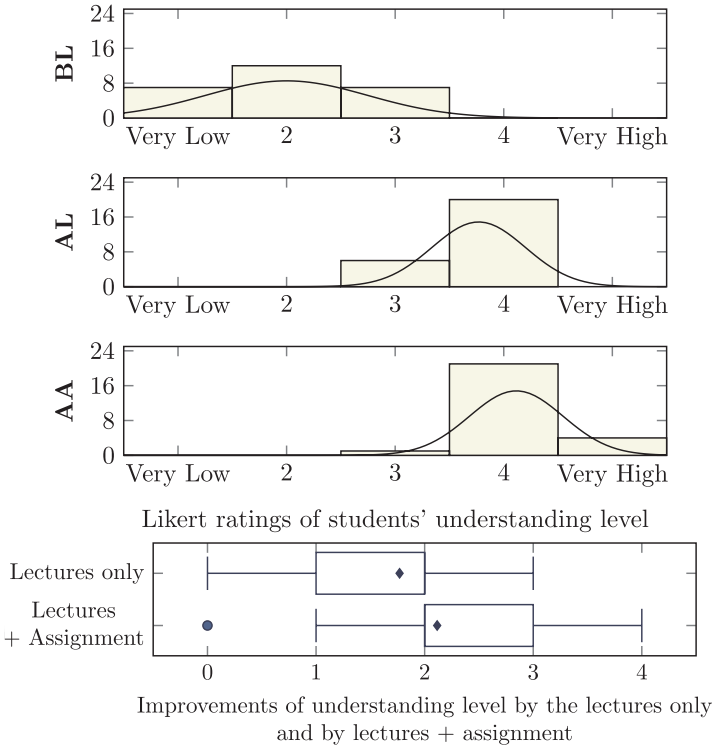


FIGURE 3. Students' understanding levels of crowdsourcing before the lectures (BL), after the lectures (AL), and after assignment (AA), and box plots of the improvements.

Student Learning and Experience

Most students used mode as a technique to get final answers, since only a few asked additional questions to differentiate the responses. Also, the overall quality of the responses was high. This made it unnecessary for them to write their own code to obtain their final answers.

We asked the students about their understanding of crowdsourcing before the lectures, after the lectures, and after this assignment on a one-to-five Likert scale where one corresponds to Know little or nothing and five to Expert. Lectures and this assignment improved their understanding levels of crowdsourcing (from 2.00 to 3.77, then to 4.12), with significance of $p \leq 0.01$ and 0.05, respectively, in the Wilcoxon signed-rank test. Figure 3 shows the distribution of students' Likert ratings in the three stages, as well as the box plots of their improvements by the lectures and overall. The curves show the normal distributions that match the means and standard deviations.

Even students who did not report an improvement in understanding postassignment accepted that the assignment was valuable in learning crowdsourcing. In fact, more than 80% of the participating students agreed that this assignment was essential to the teaching of crowdsourcing. For example, one student reported: "this assignment was very fundamental to learning the basics of crowdsourcing. It taught how to prepare a task, launch a task, receive responses, and analyze the responses." We think the lectures covered the basics of crowdsourcing well, and this assignment was a great supplement to them.

Most students needed greater detail in instructions than we provided, especially about MTurk Sandbox. Most participants reflected that their experience was "fun" or "pretty good" and that the amount of work they had to do as workers was appropriate. A few students indicated that they would have preferred an additional one or two days to complete the assignment.

Opportunities for Improvement

During the lectures, we discussed the importance of participation motivation and quality-improving strategies in crowdsourcing projects. However, in this assignment, most students asked only one straightforward question for each song in their projects, i.e., "what is the ID of this song?" Only a few students implemented additional questions to improve the quality of the responses, such as "how confident are you about your answer?" In the post-survey, almost all students realized that it was important to employ strategies to improve response quality, and more than two-thirds of them would ask different or additional questions were they to do this assignment again. We should emphasize quality-improving measures more intensively in the instructions, possibly by making them part of the grade.

The initial assignment of workers was highly imbalanced with some requesters' projects obtaining more than twenty workers and some fewer than five. We traced this problem to our placing creating projects and signing up for work in the same work segment (same deadline). Those students who wanted to complete the segment early found only a few tasks available when they signed up and decided to ignore our guideline about balancing the work assignments. Therefore, the early tasks ended up with more than twice the target number of workers. Simply breaking the project creation (as requesters) and signing up (as workers) steps could have averted this problem. We recovered by offering extra credit for additional work so that each task would receive at least ten workers. Consequently, some projects had more than ten workers, on average, those who worked on exactly ten projects received less than $10 overall.

Since we gave points for accuracy, some students applied their own expertise to identify a song. This is against the spirit of crowdsourcing. In future iterations, we will explicitly limit computed results to be based on workers' answers.

CONCLUSIONS AND RECOMMENDATIONS

This paper illustrates an approach for teaching crowdsourcing in a college course. The assignment led to good results in terms of demonstrating the value of crowdsourcing to students, improving their knowledge, and keeping them interested.

Although more than half of the participants thought that MTurk Sandbox was necessary, some thought its complexity was unnecessary because this assignment did not involve a public crowd. Workers drawn from a public crowd may be less reliable than those working on a shared task.[15] Some suggested using a public crowd to enhance realism. It would be straightforward to modify the instructions. However, a public crowd would introduce challenges such as stricter qualification tests and arranging for payments—which is administratively nontrivial in a university. We cannot impose an expense upon the students nor can we provide the funds.

Our assignment was simplified by providing high-quality snippets along with a list of possible answers. Future assignments on this theme can easily introduce interesting tasks that require human insight and adjust the level of difficulty. Some students suggested vision tasks, such as celebrity recognition from images. In addition, future assignments can emphasize parts of crowdsourcing that this assignment simplified, such as 1) coming up with a strategy for selecting workers, 2) dynamically determining the number of workers needed for a certain task, and 3) dealing with greater numbers of wrong responses.

The post-survey was filled by the students after all aspects of this assignment, including the grading, had concluded, and only self-assessments were conducted, which might have biases on students' actual understanding of the topic. However, the reported relative increases in students' understanding levels can, in fact, reflect their acknowledgment and validation of the effect of our teaching. In future offerings, rigorous tests in different stages may reflect students' improvement with greater validity.

Practice on crowdsourcing is necessary to learning it. Based on our exploratory attempt of a crowdsource-practicing assignment, we came to find some pointers that can help instructors give students better experience while achieving their course learning objectives.

- Amplify the weight of the real-life platform and give more instructions on its usage.
- Explore different tasks to solve that can

emphasize the power of crowdsourcing.
- Select the targeted crowd, public, or the class, based on your learning goals, task design, and allotted time and funds.
- Take into account the fact that students progress at different paces in a project that needs their collaborative work.

REFERENCES

1. R. Gershman, "Crowdsourcing: An old idea amplified by modern technology," Mar. 2016. [Online]. Available: http://www.onespace.com/blog/2016/03/crowdsourcing-old-ideaamplified-by-technology/. Accessed on: Sep. 15, 2017.
2. J. Vassileva, "Motivating participation in social computing applications: A user modeling perspective," *User Model. User-Adapt. Interact.*, vol. 22, no. 1, pp. 177–201, Apr. 2012.
3. C. Davidson, "How to crowdsource grading," Jul. 2009. [Online]. Available: https://www.hastac.org/blogs/cathy-davidson/2009/07/ 26/how-crowdsource-grading. Accessed on: Sep. 15, 2017.
4. P. M. Sadler and E. Good, "The impact of self- and peer-grading on student learning," *Educ. Assessment*, vol. 11, no. 1, pp. 1–31, 2006.
5. J. P. Bigham, C. Kulkarni, and W. S. Lasecki, "Crowdsourcing and crowd work," in *Proc. 2017 CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, 2017, pp. 1186–1189.
6. J. K. Goodman, C. E. Cryder, and A. Cheema, "Data collection in a flat world: The strengths and weaknesses of mechanical Turk samples," *J. Behav. Decis. Making*, vol. 26, no. 3, pp. 213–224, 2013.
7. A. J. Berinsky, G. A. Huber, and G. S. Lenz, "Using mechanical Turk as a subject recruitment tool for experimental research," *Political Anal.*, vol. 20, pp. 351–68, 2011.
8. J. S. Julia, "Identification of versions of the same musical composition by processing audio descriptions," *Ph.D. dissertation*, Univ. Pompeu Fabra, Barcelona, Spain, 2011.
9. C.-C. Wang, J.-S. Roger, and J. W. Wang, "An improved query by singing/humming system using melody and lyrics information," in *Proc. 11th Int. Soc. Music Inf. Retrieval Conf.*, Utrecht, The Netherlands, 2010, pp. 45–50.
10. A. L.-C. Wang, "An industrial strength audio search algorithm," in *Proc. 4th Int. Conf. Music Inf. Retrieval*, 2003, pp. 7–13.
11. T. Bertin-Mahieux and D. P. Ellis, "Large-scale cover song recognition using the 2D Fourier transform magnitude," in *Proc. 13th Int. Soc. Music Inf. Retrieval Conf.*, 2012, pp. 241–246.
12. V. Kharat, K. Thakare, and K. Sadafale, "A survey on query by singing/humming," *Int. J. Comput. Appl.*, vol. 111, no. 14, pp. 39–42, Feb. 2015.
13. J. Yang, J. Liu, and W.-Q. Zhang, "A fast query by humming system based on notes," in *Proc. INTERSPEECH*, 2010, pp. 2898–2901.
14. F. Kamoun, D. Alhadidi, and Z. Maamar, "Weaving risk identification into crowdsourcing lifecycle," *Procedia Comput. Sci.*, vol. 56, pp. 41–48, 2015.
15. J. P. Bigham, M. S. Bernstein, and E. Adar, "Human-computer interaction and collective intelligence," in *Collective Intelligence Handbook*. Cambridge, MA, USA: MIT Press, 2014.

HUI GUO is a Ph.D. student in computer science at North Carolina State University. His research interests include multiagent systems, natural language processing, text mining, and crowdsourcing. He received the M.S. degree in computer science from East Carolina University, and the B.S. degree from Tsinghua University. Contact him at hguo5@ncsu.edu.

NIRAV AJMERI is a Ph.D. student in computer science at North Carolina State University. His research interests include software engineering and multiagent systems with a focus on security and privacy. He received the B.E. degree in computer engineering from Sardar Vallabhbhai Patel Institute of Technology, Gujarat University. Contact him at najmeri@ncsu.edu.

MUNINDAR P. SINGH is an Alumni Distinguished Graduate Professor in computer science and a Co-Director of the Science of Security Lablet, North Carolina State University. His research interests include the engineering and governance of sociotechnical systems. He is an IEEE Fellow, an AAAI Fellow, a former Editor-in-Chief for the IEEE Internet Computing, and the current Editor-in-Chief for *ACM Transactions on Internet Technology*. Contact him at singh@ncsu.edu.

EDITOR: Erik P. DeBenedictis, Sandia National Laboratories, epdeben@sandia.gov

DEPARTMENT: REBOOTING COMPUTING

This article originally appeared in
Computer
vol. 52, no. 8, 2019

Biologically Driven Artificial Intelligence

Kjell J. Hole, *Simula UiB*

Subutai Ahmad, *Numenta*

Artificial Intelligence (AI) based on the computational principles of the brain can overcome current shortcomings and lead to human-level AI.

Artificial intelligence (AI) is the study of techniques that allow computers to learn, reason, and act to achieve goals. Machine learning is an essential type of AI that permits machines to learn from big data sets without being explicitly programmed. At the time of this writing, the international AI community is focusing on a type of machine learning called *deep learning*, a family of statistical techniques for classifying patterns using artificial neural networks (ANNs) with many layers.[1] Although deep learning has led to substantial advances in image and speech recognition, language translation, and game playing, we argue that the AI community should focus less on methods using nonbiological ANNs and more on the computational principles of the brain to create human-level or general AI with a performance close to humans at most cognitive tasks of interest.[2,3]

HUMAN VERSUS ARTIFICIAL INTELLIGENCE

Human intelligence is the brain's ability to predict sensations and events, learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate the environment. The outermost layer of the brain, the neocortex, controls cognitive functions (see "The Neocortex"). Human cognition depends on the tight connections between the brain and the body. The neocortex not only integrates sensory processing and generates motor commands, but it also uses the same sensorimotor mechanisms for high-level cognition. In other words, the structures and functions of the brain and the body form the human thought processes.

Current AI consists of computer programs that process input data from the environment and generate output data. AI researchers have mostly ignored studies of the neocortex and instead focused on mathematical and logical methods to develop AI. Because the brain is a product of evolution where newer components, including the neocortex, must rely on the functionality of the older parts, there is a widespread belief that it is possible to develop general AI by starting from scratch. AI researchers have also emphasized the limitations of the brain: the skull restricts the brain's physical size, access to energy from the body is limited, and the speed of the brain's neural circuits is slow compared to modern computers.

CURRENT LIMITATIONS OF AI

Even though it may be theoretically possible to create superior AI systems that are not biologically inspired, we are far from the goal of creating AI at the level of human intelligence. Most AI learning algorithms, particularly deep learning algorithms, are greedy, brittle, rigid, and opaque.[2] The algorithms are

Digital Object Identifier 10.1109/MC.2019.2917455
Date of publication: 30 July 2019

 Published by the IEEE Computer Society

- greedy because they demand big data sets to learn
- brittle because they frequently fail when confronted with a mildly different scenario than that in the training set
- rigid because they cannot keep adapting after initial training
- opaque because the internal representations make it challenging to interpret their decisions.

Although these shortcomings are all serious, the core problem is that all AI systems are shallow because they lack abstract reasoning abilities and possess no common sense about the world.

The most popular AI methods today use ANNs. The brain was the original inspiration for ANNs, but this is no longer the case; the knowledge behind the ANNs' brain cells is outdated. We now know that biological neurons have multiple physical states and biological networks have far more sophisticated functionality than those in ANNs. Furthermore, most AI systems require separate, offline training periods, whereas learning in the neocortex occurs continuously and in real time using data streaming in from all our senses. Finally, sensorimotor processing is not deeply integrated into most AI methods. It is therefore unclear whether current AI approaches can lead to general human-level intelligence.

BIOLOGICALLY BASED AI

The limited progress toward general AI after decades of research strongly indicates that few paths lead to human-level intelligence and many paths lead nowhere. Not surprisingly, without guidance, it is challenging to discover the right approach in a vast space of algorithms containing very few solutions. Instead, the community should focus on the only example we have of intelligence, namely, the brain and, especially, the neocortex.

There has been much effort in neuroscience to reverse engineer mammalian brains, especially to understand what the neocortex does and how it works. Substantial reverse engineering efforts include the U.S. government program Machine Intelligence from Cortical Networks, the Swiss Blue Brain Project, and the Human Brain Project funded mainly by the European Union. The research lab Numenta, led by Jeff Hawkins, has shared a theoretical framework, called *Hierarchical Temporal Memory* (*HTM*), that describes the computational principles of the neocortex.[4–6] Numenta has made a sustained effort to make HTM understandable to people without a neuroscience background (see research papers and tutorial videos at numenta.com).

One of the findings of neuroscience is that the neocortex consists of a repeating circuit, known as a *cortical column*, that creates our perceptions, language, and high-level thoughts. HTM theory models several components of this circuit that, together, have led to a new functional interpretation of the neocortex. We summarize these findings and the differences between biologically based frameworks and ANN-based systems in the following sections.[4–6]

Realistic neuron model

AI models neurons as a simple function operating on a single set of connections. Biological neurons, modeled by the HTM neuron, have segregated feedforward, lateral, and feedback connections that detect multiple independent patterns of neuronal activity. Whereas the ANN neuron is either active or inactive, the HTM neuron has more states. The HTM neuron goes into a predictive state when it is soon likely to become active based on lateral and feedback context. The ability to recognize contextual patterns and predict future states form the basis for a predictive sequence memory that allows HTM to process multiple hypotheses about the world simultaneously.

THE NEOCORTEX

The neocortex is an intensely folded sheet with a thickness of approximately 2.5 mm. The folds substantially increase its surface area. When laid out flat, the neocortex is the size of a formal dinner napkin. It constitutes roughly 75% of the brain's volume and contains 30 billion cells called *neurons* [see Figure S1(a)]. A typical neuron has one tail-like axon and several tree-like extensions called *dendrites*. When a cell fires, an electrochemical pulse or spike travels down the axon to its terminals. A signal jumps from an axon terminal to the receptors on a dendrite of another neuron. The axon terminal, the receptors, and the cleft between them constitute a synapse (not shown). The axon terminal releases neurotransmitters into the synaptic cleft to signal the dendrite. The neuron is thus a signaling system in which the axon is the transmitter, the dendrites are the receivers, and the synapses are the connectors between them. A neuron has between 5,000 and 20,000 synapses.

The neocortex consists of cortical columns as depicted in Figure S1(b). The columns have six layers, five of which contain neurons. The thickness of a layer varies throughout the neocortex, but the basic structure is similar. As a result, the neocortex has both a vertical columnar and horizontal laminar organization. Cortical columns are grouped into regions connected through bundles of nerve fibers and organized in hierarchies. The regions realizing vision, language, and touch all have highly similar structures and thus operate according to the same principles. The sensory input determines a region's purpose. Some regions receive inputs directly from the sensory organs, while others receive them from other regions. The neocortex generates body movements to change the sensory inputs and learn quickly about the world. Learning occurs by growing and removing synapses, that is, by changing the network itself.

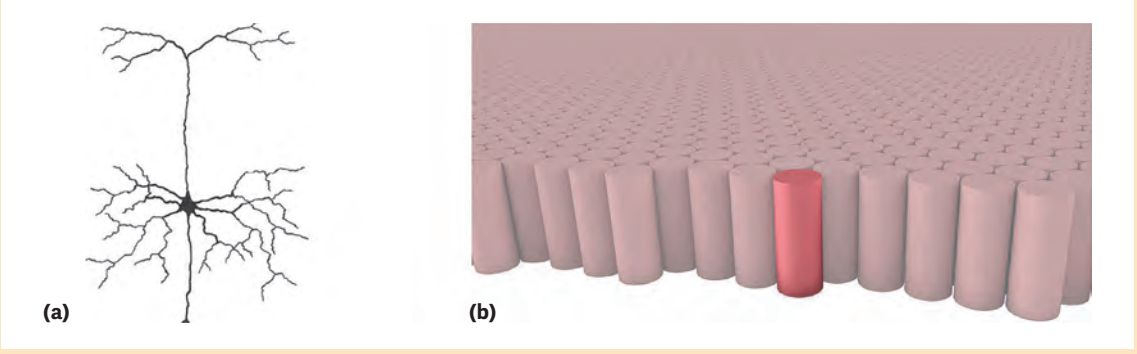


FIGURE S1. (a) A biological neuron in the neocortex. The axon is the lowermost vertical connection. (b) A flattened version of the neocortex consisting of cortical columns, all with a similar structure.

Sparse data representations

An ANN uses dense scalar vectors to represent data. Such representations are vulnerable to errors and noise, and dense vector-matrix multiplications consume significant power. The brain uses sparse data representations where only a small percentage of the neurons is active, and each neuron is only connected to a fraction of other neurons. Sparse representations allow HTM to represent and process patterns of neuronal activity in a brain-like manner that is robust to changes and faults and consume significantly less power.

Sensorimotor integration

The brain models objects and environments by integrating sensations and movement-derived location signals. These object models are continuously updated as sensors move to create movement-invariant models. HTM uses such models to make predictions, detect anomalies, and thus learn about the world. It is unclear how to achieve the same results with ANNs.

Continuous online learning

To train most ANN systems, we must first assemble

big training sets and then run learning algorithms offline. HTM learning occurs in real time using streaming data from sensors. Learning in the neocortex, and in HTM, dynamically rewires the connections between neurons to record information. Since the Hebbian-like learning rules are local to each HTM neuron, learning in the neocortex can scale to train massive systems.

The Thousand Brains Theory of Intelligence

Connectivity between cortical columns is strongly nonhierarchical, with numerous lateral and feedback pathways. HTM theory states that instead of learning one big model of the world, each cortical column learns a separate movement-invariant model. Communication between cortical columns cuts across hierarchy and sensory modalities and rapidly resolves uncertainty. The neocortex thus learns thousands of models that operate in a massively parallel fashion. This thousand brains theory of intelligence is built on a single universal learning algorithm embodied by the cortical column. The algorithm learns objects, behaviors, and concepts by representing compositional structures, learning through movement, and integrating knowledge across different senses. There is no comparable theory based on ANNs.

FURTHER RESEARCH

Although deep learning can outperform humans in isolated domains characterized by fixed rules and little need for contextual knowledge,[1] we require novel ideas to create general AI. The biological path toward human-level AI, particularly the principles discussed previously, deserves more attention from the AI community because biologically constrained AI avoids the limitations of the approaches based on ANNs. More computer scientists should study neuroscience and use the computational principles of the brain to develop AI. Students and budding researchers with a desire to contribute to the development of AI should also study neuroscience. Finally, we need increased cooperation with neuroscientists, given how hard it is to select relevant results from the vast neuroscience literature.

There is a rich road map for future research into biologically constrained AI. Additional reverse engineering is required to understand the organization of the neocortical regions and how they collaborate to create intelligent behavior. It is also necessary to reverse engineer other parts of the brain, such as the thalamus, which is intimately involved in the communication between the neocortical regions.

It will take effort to follow the biological path to general AI, but repeated improvements along the way could gradually change society for the better. Initially, we could build specialized brains to improve cybersecurity or solve difficult problems in mathematics, physics, or medicine. Eventually, we could create autonomous intelligent systems to carry out tedious or dangerous work.

REFERENCES

1. T. J. Sejnowski, *The Deep Learning Revolution*. Cambridge, MA: MIT Press, 2018.
2. G. Marcus, *Deep learning: A critical appraisal*, 2018. [Online]. Available: https://arxiv.org/abs/1801.00631.
3. D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.
4. J. Hawkins and S. Ahmad, "Why neurons have thousands of synapses, a theory of sequence memory in neocortex," *Front. Neural Circuits*, vol. 10, March 2016. [Online]. Available: https://doi.org/10.3389/fncir.2016.00023
5. J. Hawkins, S. Ahmad, and Y. Cui, "A theory of how columns in the neocortex enable learning the structure of the world," *Front. Neural Circuits*, vol. 11, Oct. 2017. [Online]. Available: https://doi.org/10.3389/fncir.2017.00081
6. J. Hawkins, M. Lewis, M. Klukas, S. Purdy, and S. Ahmad, "A framework for intelligence and cortical function based on grid cells in the neocortex," *Front. Neural Circuits*, vol. 12, Jan. 2019. [Online]. Available: https://doi.org/10.3389/fncir.2018.00121

KJELL J. HOLE is the director of the research center Simula UiB, Bergen, Norway. Contact him at hole@simula.no.

SUBUTAI AHMAD is the vice president of research at Numenta, Redwood City, California. Contact him at sahmad@numenta.com.

NUS National University of Singapore

Faculty Positions in Computer Science

The Department of Computer Science at the National University of Singapore (NUS) invites applications for tenure-track and educator-track positions in all areas of computer science.

The Department is looking for candidates for all levels of tenured and tenure-track positions in any area of computer science. Candidates for Assistant Professor positions on the tenure track should be early in their academic careers and yet demonstrate outstanding research potential, and a strong commitment to teaching.

For Senior Lecturer and Associate Professor on the educator-track, teaching experience or relevant industry experience will be preferred. Besides relevant background and experience, we are also looking for someone with a passion for imparting the latest knowledge in computing to students in our programs

The Department enjoys ample research funding, moderate teaching loads, excellent facilities, and extensive international collaborations. We have a full range of faculty covering all major research areas in computer science and boasts a thriving PhD program that attracts the brightest students from the region and beyond. More information is available https://www.comp.nus.edu.sg/about/depts/cs/recruitment/faculty/

NUS is an equal opportunity employer that offers highly competitive salaries, and is situated in Singapore, an English-speaking cosmopolitan city that is a melting pot of many cultures, both the east and the west. Singapore offers high-quality education and healthcare at all levels, as well as very low tax rates.

Application Details:

Submit the following documents (in a single PDF) online via: https://faces.comp.nus.edu.sg

- A cover letter that indicates the position applied for and the main research interests
- Curriculum Vitae
- A teaching statement
- A research statement (optional but encouraged for educator-track)

Provide the contact information of 3 referees when submitting your online application, or, arrange for at least 3 references to be sent directly to csrec@comp.nus.edu.sg

To ensure maximal consideration, please submit your application by 18 December 2020.

If you have further enquiries, please contact the Search Committee Chair, Joxan Jaffar, at csrec@comp.nus.edu.sg

TECHNOLOGY

Oracle America, Inc.

has openings for the following positions **(all levels/types)** in San Mateo County, including Redwood Shores, CA and San Bruno, CA; Alameda County, including Pleasanton, CA; San Francisco, CA; Santa Clara County, including Santa Clara, CA and San Jose, CA; and other locations in the San Francisco Bay Area. Some positions may allow for telecommuting.

Hardware Developers (HWD1020): Evaluate reliability of materials, properties and techniques used in production; plan, design and develop electronic parts, components, integrated circuitry, mechanical systems, equipment and packaging, optical systems and/or DSP systems.

Product Managers (PM1020): Participate in all software and/or hardware product development life cycle activities. Move software products through the software product development cycle from design and development to implementation, testing, and/or marketing.

Software Developers (SWD1020): Design, develop, troubleshoot and/or test/QA software.

Applications Developers (APD1020): Analyze, design, develop, troubleshoot and debug software programs for commercial or end user applications. Write code, complete programming and perform testing and debugging of applications.

Programmer Analysts (PA1020): Analyze user requirements to develop, implement, and/or support Oracle's global infrastructure.

Technical Analysts (TA1020): Deliver solutions to the Oracle customer base while serving as an advocate for customer needs. Offer strategic technical support to assure the highest level of customer satisfaction.

Consultants (TCONS1020): Analyze requirements and deliver functional and technical solutions. Implement products and technologies to meet post-sale customer needs. Travel to various unanticipated sites throughout the U.S. required.

Sales Consultants (TSC1020): Provide presales technical/functional support to prospective customers. Design, validate and present software solutions to include product concepts and future direction. Travel to various unanticipated sites throughout the U.S. required.

Software Developers (TSWD1020): Design, develop, troubleshoot and/or test/QA software. Travel to various unanticipated sites throughout the U.S. required.

Applications Developers (TAPD1020): Analyze, design, develop, troubleshoot and debug software programs for commercial or end user applications. Write code, complete programming and perform testing and debugging of applications. Travel to various unanticipated sites throughout the U.S. required.

Product Managers (TPM1020): Participate in all software and/or hardware product development life cycle activities. Move software products through the software product development cycle from design and development to implementation, testing, and/or marketing. Travel to various unanticipated sites throughout the U.S. required.

Submit resume to applicant_us@oracle.com. Must include job#. Oracle supports workforce diversity.

Conference Calendar

IEEE Computer Society conferences are valuable forums for learning on broad and dynamically shifting topics from within the computing profession. With over 200 conferences featuring leading experts and thought leaders, we have an event that is right for you. Questions? Contact conferences@computer.org.

NOVEMBER

2 November
- Blockchain (IEEE Int'l Conf. on Blockchain), virtual

4 November
- CONISOFT (Int'l Conf. in Software Engineering Research and Innovation), virtual

6 November
- CCEM (IEEE Int'l Conf. on Cloud Computing in Emerging Markets), virtual
- SmartCloud (IEEE Int'l Conf. on Smart Cloud), virtual

9 November
- ICTAI (IEEE Int'l Conf. on Tools with Artificial Intelligence), virtual
- IRC (IEEE Int'l Conf. on Robotic Computing), virtual
- ISMAR (IEEE Int'l Symposium on Mixed and Augmented Reality), virtual
- ISMVL (IEEE Int'l Symposium on Multiple-Valued Logic), virtual

11 November
- SEC (IEEE/ACM Symposium on Edge Computing), virtual

15 November
- SC, virtual

16 November
- FOCS (IEEE Symposium on Foundations of Computer Science), Durham, USA
- LCN (IEEE Conf. on Local Computer Networks), virtual

17 November
- ICDM (IEEE Int'l Conf. on Data Mining), virtual

26 November
- NCA (IEEE Int'l Symposium on Network Computing and Applications), virtual

29 November
- ICDCS (IEEE Int'l Conf. on Distributed Computing Systems), virtual

30 November
- ICHI (IEEE Int'l Conf. on Healthcare Informatics), Oldenburg, Germany

DECEMBER

1 December
- ICRC (IEEE Int'l Conf. on Rebooting Computing), virtual

2 December
- CSDE (IEEE Asia-Pacific Conf. on Computer Science and Data Eng.), Gold Coast, Australia
- IEEE InTech (A Forum on the Response and Resiliency to COVID-19), virtual
- ISM (IEEE Int'l Symposium on Multimedia), virtual

6 December
- HOST (IEEE Int'l Symposium on Hardware-Oriented Security and Trust), virtual

7 December
- ASONAM (IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining), virtual
- BDCAT (IEEE/ACM Int'l Conf. on Big Data Computing, Applications and Technologies), virtual
- UCC (IEEE/ACM Int'l Conf. on Utility and Cloud Computing), virtual

9 December
- CC (IEEE Int'l Conf. on Conversational Computing), virtual
- AIKE (IEEE Int'l Conf. on Artificial Intelligence and Knowledge Eng.), virtual

10 December
- BigData (IEEE Int'l Conf. on Big Data), virtual

14 December
- AIVR (IEEE Int'l Conf. on Artificial Intelligence and Virtual Reality), virtual
- CloudCom (IEEE Int'l Conf. on Cloud Computing Technology and Science), Bangkok, Thailand

- HPCC (IEEE Int'l Conf. on High Performance Computing and Communications), virtual

16 December

- BIBM (IEEE Int'l Conf. on Bioinformatics and Biomedicine), virtual
- HiPC (IEEE Int'l Conf. on High-Performance Computing, Data, and Analytics), virtual

29 December

- BigDataSE (IEEE Int'l Conf. on Big Data Science and Eng.), Guangzhou, China
- EUC (IEEE Int'l Conf. on Embedded and Ubiquitous Computing), Guangzhou, China
- TrustCom (IEEE Int'l Conf. on Trust, Security and Privacy in Computing and Communications), Guangzhou, China

2021

JANUARY

5 January

- WACV (IEEE Winter Conf. on Applications of Computer Vision), virtual

10 January

- ICPR (Int'l Conf. on Pattern Recognition), Milan, Italy

17 January

- BigComp (IEEE Int'l Conf. on Big Data and Smart Computing), Bangkok, Thailand

27 January

- ICSC (IEEE Int'l Conf. on Semantic Computing), Laguna Hills, USA

FEBRUARY

27 February

- CGO (IEEE/ACM Int'l Symposium on Code Generation and Optimization), virtual
- HPCA (IEEE Int'l Symposium on High-Performance Computer Architecture), virtual

MARCH

22 March

- PerCom (IEEE Int'l Conf. on Pervasive Computing and Communications), Kassel, Germany
- MIPR (IEEE Int'l Conf. on Multimedia Information Processing and Retrieval), Tokyo, Japan

27 March

- IEEE VR (IEEE Conf. on Virtual Reality and 3D User Interfaces), Lisbon, Portugal

APRIL

12 April

- ICST (IEEE Conf. on Software Testing, Verification and Validation), virtual

19 April

- ICDE (IEEE Int'l Conf. on Data Engineering), Chania, Greece

21 April

- SELSE (IEEE Workshop on Silicon Errors in Logic - System Effects), Los Angeles, USA

MAY

10 May

- CCGrid (IEEE/ACM Int'l Symposium on Cluster, Cloud and Internet Computing), Melbourne, Australia

17 May

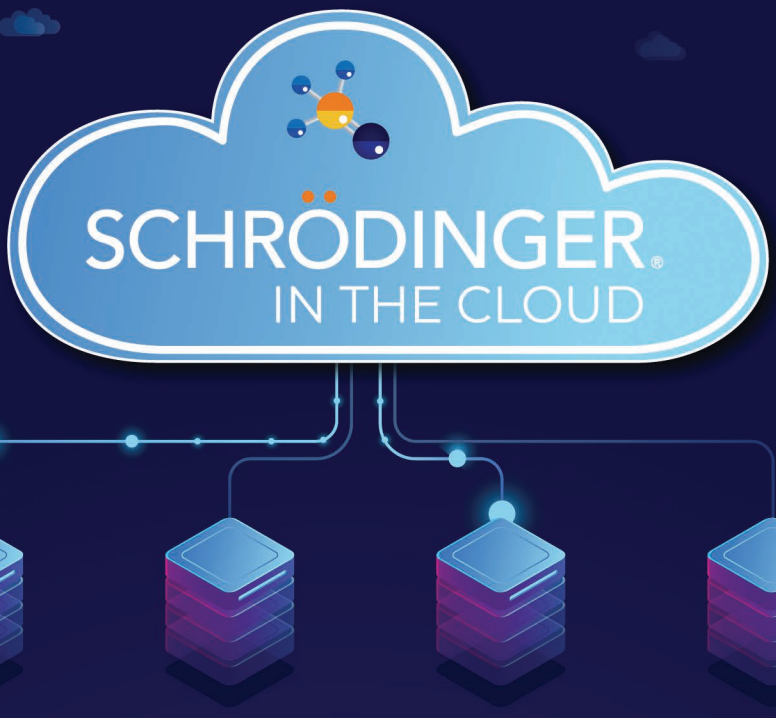
- IPDPS (IEEE Int'l Parallel and Distributed Processing Symposium), Portland, Oregon, USA

23 May

- SP (IEEE Symposium on Security and Privacy), San Francisco, USA

25 May

- ISMVL (IEEE Int'l Symposium on Multiple-Valued Logic), Nur-Sultan, Kazakhstan



High-performance Computing with Schrödinger's leading physics-based software in the cloud

Use Schrödinger's GUI Maestro directly in-browser or submit jobs and amplify the backend computing resources you need, when you need them, with the auto-scaling Virtual Clusters of *Schrödinger in the Cloud*

RESULTS IN MINUTES

REDUCE COSTS

SCALE AUTOMATICALLY

For more information visit
schrodinger.com/cloudcomputing