

COMMUNICATIONS

CACM.ACM.ORG

OF THE

ACM

10/2021 VOL.64 NO.10



Human Detection of Machine-Manipulated Media

Trustworthy AI

Let's Fix the Internet, Not the Tech Giants

The Role of Professional Certification

Flatter Chips

Association for
Computing Machinery

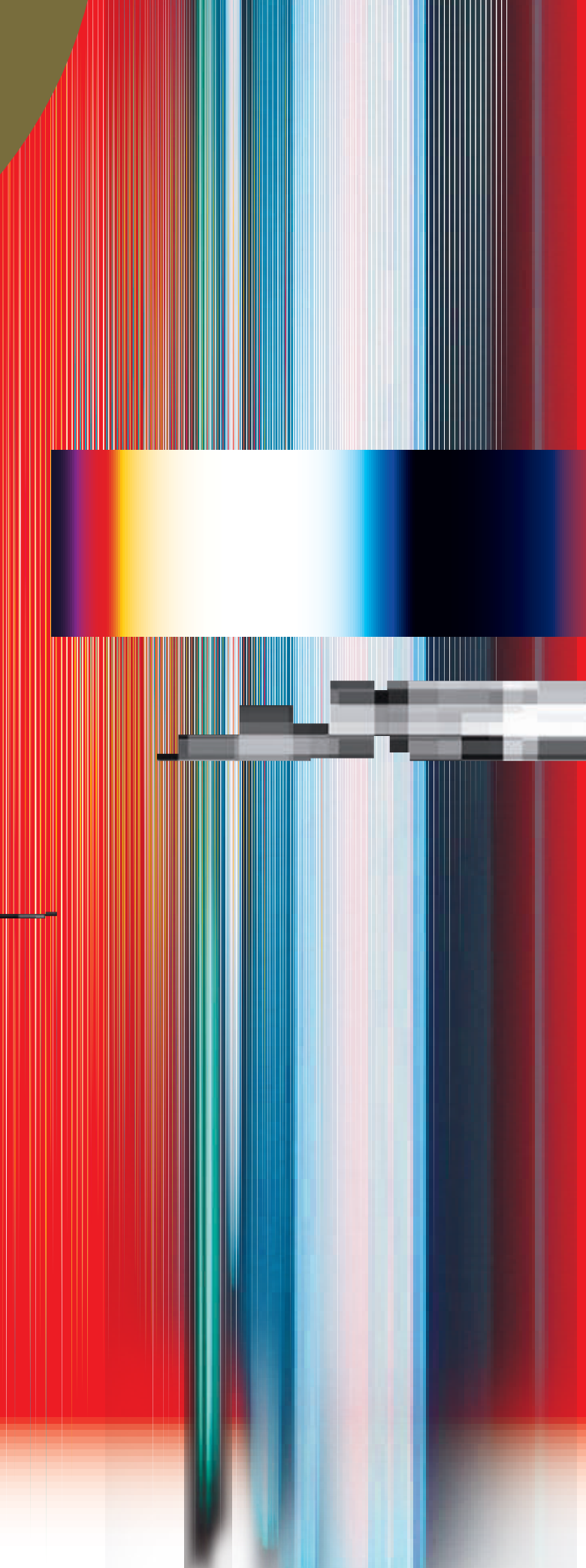
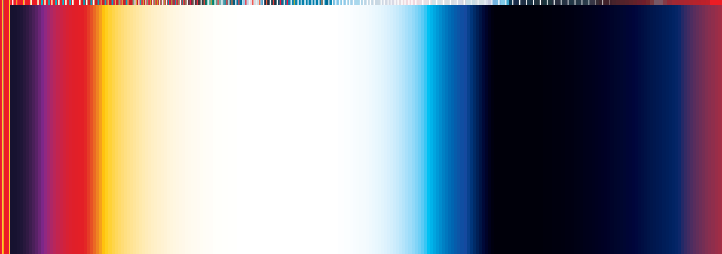
acm



SIGGRAPH ASIA 2021 TOKYO

CONFERENCE 14 - 17 DECEMBER 2021
EXHIBITION 15 - 17 DECEMBER 2021
TOKYO INTERNATIONAL FORUM, JAPAN

sa2021.siggraph.org



ONE GIFT

WORLD-CHANGING IMPACT

25 MILLION
SMALL STEPS TOWARD

ADVANCING KNOWLEDGE
REVOLUTIONIZING EDUCATION
TRANSFORMING SOCIETY

115+
FACULTY

1,700+
UNDERGRADUATE
STUDENTS

1,000+
GRADUATE
STUDENTS

| **Aiming for the Pinnacle of Excellence at Scale** |

Learn about the Elmore family's contribution at:

PURDUE.LINK/ELMORE



Elmore Family School of Electrical
and Computer Engineering

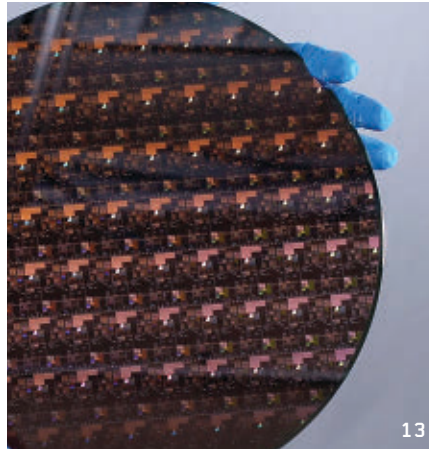
Departments

- 5 **Cerf's Up**
The Future of Text Redux
By Vinton G. Cerf
-
- 6 **BLOG@CACM**
New Life for Cordless Communication, Old Regrets for Software Projects
Andrei Sukhov and Igor Sorokin ponder the potential benefits of DECT to the Internet of Things, while Doug Meil considers how software engineers should reflect on their accomplishments.
-
- 94 **Careers**

Last Byte

- 96 **Upstart Puzzles**
Randomower
By Dennis Shasha

News



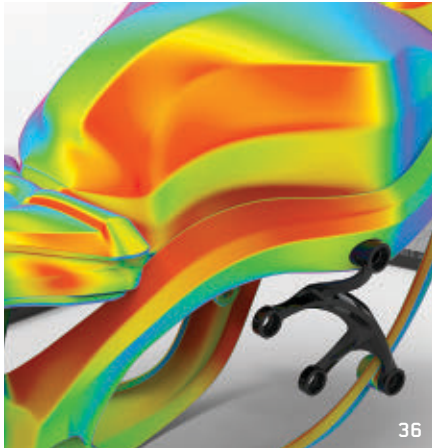
- 8 **Flatter Chips**
Two-dimensional materials—graphene and its cousins—could enable better integrated circuits.
By Don Monroe
-
- 11 **Algorithmic Poverty**
Algorithms can have a devastating impact on people's lives, especially if they already are struggling economically.
By Keith Kirkpatrick
-
- 13 **A Switch in Time**
The introduction of the nanosheet transistor is just one step in the continuing attempt to stick to the spirit of Moore's Law.
By Chris Edwards

Viewpoints

- 16 **Technology Strategy and Management Section 230 and a Tragedy of the Commons**
The dilemma of social media platforms.
By Michael A. Cusumano
-
- 19 **Broadening Participation**
Broadening Participation by Teaching Accessibility
Strategies for incorporating accessibility into computing education.
By Kendra Walther and Richard E. Ladner
-
- 22 **Global Computing**
Remaining Connected Throughout Design
Applying the unique experiences of designing technologies for vulnerable communities.
By Hannah Thinyane
-
- 25 **Kode Vicious**
Divide and Conquer
The use and limits of bisection.
By George V. Neville-Neil
-
- 26 **Viewpoint**
Competitive Compatibility: Let's Fix the Internet, Not the Tech Giants
Seeking to make Big Tech less central to the Internet.
By Cory Doctorow
-
- 30 **Viewpoint**
AI Futures: Fact and Fantasy
Three books offer varied perspectives on the ascendancy of artificial intelligence.
By Devdatt Dubhashi



Practice



36

32 **Software Development in Disruptive Times**

Creating a software solution with fast decision capability, agile project management, and extreme low-code technology.
By João Varajão

36 **A New Era for Mechanical CAD**

Time to move forward from decades-old design.
By Jessie Frazelle

Q Articles' development led by acmqueue.queue.acm.org

About the Cover:



Neil Armstrong's photo of fellow astronaut Buzz Aldrin on the moon's surface in July 1969 ranks among the most famous images of all time. But wait, something's different ... how did *that* happen? This month's cover story (p. 40) explores the power of AI to manipulate media and how good we humans are at detecting it. Cover image composite by Andrij Borys Associates.

IMAGES IN COVER COLLAGE: Apollo 11 photo courtesy of NASA (Public Domain); UAV helicopter © 2021 by Pedro Jordano (CC BY-NC-SA 4.0).

Contributed Articles



48

40 **Human Detection of Machine-Manipulated Media**

Technologies for manipulating and faking online media may outpace people's ability to tell the difference.
By Matthew Groh, Ziv Epstein, Nick Obradovich, Manuel Cebrian, and Iyad Rahwan



Watch the authors discuss this article in the exclusive *Communications* video. <https://cacm.acm.org/videos/machine-manipulated-media>

48 **Six Reasons Why Virtual Reality Is a Game-Changing Computing and Communication Platform for Organizations**

Beyond the pandemic, organizations need to recognize what digital assets, interactions, and communication processes reap the most benefits from virtual reality.
By Osku Torro, Henri Jalo, and Henri Pirkkalainen

56 **The Role of Professional Certifications in Computer Occupations**

Experienced and aspiring computing professionals need to manage their qualifications according to current market needs.
By Mark Tannian and Willie Coston

Review Articles

64 **Trustworthy AI**
The pursuit of responsible AI raises the ante on both the trustworthy computing and formal methods communities.

By Jeannette M. Wing



Watch the author discuss this article in the exclusive *Communications* video. <https://cacm.acm.org/videos/trustworthy-ai>

Research Highlights

74 **Technical Perspective**
Liquid Testing Using Built-in Phone Sensors
By Tam Vu

75 **Liquid Testing with Your Smartphone**
By Shichao Yue and Dina Katabi

84 **Technical Perspective**
The Real-World Dilemma of Security and Privacy by Design
By Ahmad-Reza Sadeghi

85 **Securing the Wireless Emergency Alerts System**
By Jihoon Lee, Gyuhong Lee, Jinsung Lee, Youngbin Im, Max Hollingsworth, Eric Wustrow, Dirk Grunwald, and Sangtae Ha



ACM, the world's largest educational and scientific computing society, delivers resources that advance computing as a science and profession. ACM provides the computing field's premier Digital Library and serves its members and the computing profession with leading-edge publications, conferences, and career resources.

Executive Director and CEO

Vicki L. Hanson

Deputy Executive Director and COO

Patricia Ryan

Director, Office of Information Systems

Wayne Graves

Director, Office of Financial Services

James Schembari

Director, Office of SIG Services

Donna Cappel

Director, Office of Publications

Scott E. Delman

ACM COUNCIL

President

Gabriele Kotsis

Vice-President

Joan Feigenbaum

Secretary/Treasurer

Elisa Bertino

Past President

Cherri M. Pancake

Chair, SGB Board

Jeff Jortner

Co-Chairs, Publications Board

Joseph Konstan and Divesh Srivastava

Members-at-Large

Nancy M. Amato; Tom Crick; Susan Dumais; Mehran Sahami; Alejandro Saucedo

SGB Council Representatives

Sarita Adve and Jeanna Neefe Matthews

BOARD CHAIRS

Education Board

Elizabeth Hawthorne and Chris Stephenson

Practitioners Board

Terry Coatta

REGIONAL COUNCIL CHAIRS

ACM Europe Council

Chris Hankin

ACM India Council

Abhiram Ranade

ACM China Council

Wenguang Chen

PUBLICATIONS BOARD

Co-Chairs

Joseph Konstan and Divesh Srivastava

Board Members

Jonathan Aldrich; Jack Davidson; Chris Hankin; Mike Heroux; James Larus; Marc Najork; Michael L. Nelson; Holly Rushmeier; Eugene H. Spafford; Bhavani Thuraisingham; Julie R. Williamson

ACM U.S. Technology Policy Office

Adam Eisgrau

Director of Global Policy and Public Affairs

1701 Pennsylvania Ave NW, Suite 200,

Washington, DC 20006 USA

T (202) 580-6555; acmpo@acm.org

Computer Science Teachers Association

Jake Baskin

Executive Director

COMMUNICATIONS OF THE ACM

Trusted insights for computing's leading professionals.

Communications of the ACM is the leading monthly print and online magazine for the computing and information technology fields. *Communications* is recognized as the most trusted and knowledgeable source of industry information for today's computing professional. *Communications* brings its readership in-depth coverage of emerging areas of computer science, new trends in information technology, and practical applications. Industry leaders use *Communications* as a platform to present and debate various technology implications, public policies, engineering challenges, and market trends. The prestige and unmatched reputation that *Communications of the ACM* enjoys today is built upon a 50-year commitment to high-quality editorial content and a steadfast dedication to advancing the arts, sciences, and applications of information technology.

STAFF

DIRECTOR OF PUBLICATIONS

Scott E. Delman
cacm-publisher@cacm.acm.org

Executive Editor

Diane Crawford

Managing Editor

Thomas E. Lambert

Senior Editor

Ralph Raiola

Senior Editor/News

Lawrence M. Fisher

Web Editor

David Roman

Editorial Assistant

Danbi Yu

Art Director

Andrij Borys

Associate Art Director

Margaret Gray

Assistant Art Director

Mia Angelica Balaquiot

Production Manager

Bernadette Shade

Intellectual Property Rights Coordinator

Barbara Ryan

Advertising Sales Account Manager

Iliia Rodriguez

Columnists

David Anderson; Michael Cusumano; Peter J. Denning; Mark Guzdial; Thomas Haigh; Leah Hoffmann; Mari Sako; Pamela Samuelson; Marshall Van Alstyne

CONTACT POINTS

Copyright permission
permissions@hq.acm.org

Calendar items

calendar@cacm.acm.org

Change of address

acmhelp@acm.org

Letters to the Editor

letters@cacm.acm.org

REGIONAL SPECIAL SECTIONS

Co-Chairs

Jakob Rehof, Haibo Chen, and P J Narayanan

Board Members

Sherif G. Aly; Panagioti Fatourou; Chris Hankin; Sue Moon; Tao Xie; Kenjiro Taura; David Padua

WEBSITE

http://cacm.acm.org

WEB BOARD

Chair

James Landay

Board Members

Marti Hearst; Jason I. Hong; Jeff Johnson; Wendy E. MacKay

AUTHOR GUIDELINES

http://cacm.acm.org/about-communications/author-center

ACM ADVERTISING DEPARTMENT

1601 Broadway, 10th Floor
New York, NY 10019-7434 USA
T (212) 626-0686
F (212) 869-0481

Advertising Sales Account Manager

Iliia Rodriguez
ilia.rodriguez@hq.acm.org

Media Kit acmm mediasales@acm.org

EDITORIAL BOARD

EDITOR-IN-CHIEF

Andrew A. Chien
aie@cacm.acm.org

Deputy to the Editor-in-Chief

Morgan Denlow
cacm.deputy.to.eic@gmail.com

SENIOR EDITOR

Moshe Y. Vardi

NEWS

Co-Chairs

Marc Snir and Alain Chesnais

Board Members

Tom Conte; Monica Divitini; Mei Kobayashi; Rajeev Rastogi; François Sillion

VIEWPOINTS

Co-Chairs

Tim Finin; Susanne E. Hambrusch;

John Leslie King

Board Members

Virgilio Almeida; Terry Benzel; Michael L. Best; Judith Bishop; Lorrie Cranor; Boi Falting; James Gimmelmann; Mark Guzdial; Haym B. Hirsch; Anupam Joshi; Richard Ladner; Carl Landwehr; Beng Chin Ooi; Francesca Rossi; Len Shustek; Loren Terveen; Marshall Van Alstyne; Jeannette Wing; Susan J. Winter

PRACTICE

Co-Chairs

Stephen Bourne and Theo Schlossnagle

Board Members

Eric Allman; Samy Bahra; Peter Bailis; Betsy Beyer; Terry Coatta; Stuart Feldman; Nicole Forsgren; Camille Fournier; Jessie Fratzelle; Benjamin Fried; Tom Killalea; Tom Limoncelli; Kate Matsudaira; Marshall Kirk McKusick; Erik Meijer; George Neville-Neil; Jim Waldo; Meredith Whittaker

CONTRIBUTED ARTICLES

Co-Chairs

James Larus and Gail Murphy

Board Members

Robert Austin; Nathan Baker; Kim Bruce; Alan Bundy; Peter Buneman; Premkumar T. Devanbu; Jane Cleland-Huang; Yannis Ioannidis; Rebecca Isaacs; Trent Jaeger; Somesh Jha; Gal A. Kaminka; Ben C. Lee; Igor Markov; m.c. schraefel; Hannes Werthner; Ryan White; Reinhard Wilhelm; Rich Wolski

RESEARCH HIGHLIGHTS

Co-Chairs

Shriram Krishnamurthi

and Orna Kupferman

Board Members

Martin Abadi; Amr El Abbadi; Animashree Anandkumar; Sanjeev Arora; Michael Backes; Maria-Florina Balcan; Azer Bestavros; David Brooks; Stuart K. Card; Jon Crowcroft; Lieven Eeckhout; Alexei Efros; Bryan Ford; Alon Halevy; Gernot Heiser; Takeo Igarashi; Srinivasan Keshav; Sven Koenig; Ran Libeskind-Hadas; Karen Liu; Joanna McGrenere; Tim Roughgarden; Guy Steele, Jr.; Robert Williamson; Margaret H. Wright; Nicholai Zeldovich; Andreas Zeller

Association for Computing Machinery (ACM)

1601 Broadway, 10th Floor
New York, NY 10019-7434 USA
T (212) 869-7440; F (212) 869-0481

ACM Copyright Notice

Copyright © 2021 by Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or fee. Request permission to publish from permissions@hq.acm.org or fax (212) 869-0481.

For other copying of articles that carry a code at the bottom of the first or last page or screen display, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center; www.copyright.com.

Subscriptions

An annual subscription cost is included in ACM member dues of \$99 (\$40 of which is allocated to a subscription to *Communications*); for students, cost is included in \$42 dues (\$20 of which is allocated to a *Communications* subscription). A nonmember annual subscription is \$269.

ACM Media Advertising Policy

Communications of the ACM and other ACM Media publications accept advertising in both print and electronic formats. All advertising in ACM Media publications is at the discretion of ACM and is intended to provide financial support for the various activities and services for ACM members. Current advertising rates can be found by visiting <http://www.acm-media.org> or by contacting ACM Media Sales at (212) 626-0686.

Single Copies

Single copies of *Communications of the ACM* are available for purchase. Please contact acmhelp@acm.org.

COMMUNICATIONS OF THE ACM

(ISSN 0001-0782) is published monthly by ACM Media, 1601 Broadway, 10th Floor New York, NY 10019-7434 USA. Periodicals postage paid at New York, NY 10001, and other mailing offices.

POSTMASTER

Please send address changes to *Communications of the ACM* 1601 Broadway, 10th Floor New York, NY 10019-7434 USA

Printed in the USA.



Association for Computing Machinery





Vinton G. Cerf

DOI:10.1145/3483539

The Future of Text Redux

I HAVE WRITTEN ABOUT text in its digital form in the past and I would like to revisit this topic once more. J.C.R. Licklider and Douglas Engelbart were two giants who saw non-numeric possibilities in networked computing well ahead of many others. Vannevar Bush and Ted Nelson are two others who resonated with the idea of machines that assisted in the production and discovery of information. Sir Tim Berners-Lee amplified some of these ideas with the invention of the World Wide Web. Following along this path is Frode Hegland, a protégé of the late Douglas Engelbart, who has developed new tools for the production of and interaction with text. Engelbart's oNLine System (NLS)^a was a tour-de-force example of disciplined use of structure to guide the production and consumption of computer-based text. One could view content at various depths (for example, first line of each paragraph, first paragraph only, subsets of segments of text found in classic structured outlines of document) and one could accomplish major restructuring of documents with ease because the NLS understood the structure and provided ways to reference portions of documents to facilitate restructuring.

Hegland has developed three remarkable tools for text generation, viewing, and referencing that inherit some of the philosophical aspects of NLS and enrich them with more fluid ways of organizing, viewing, and generating text. His three contributions are AUTHOR, READER, and VISUAL-META.^b These three tools illustrate

the power of applying computing to text, creating lenses through which to create, consume, and reference content. Hegland's focus is less on the appearance of text, over which most text editors persevere, than on its structure and relationships among various parts of a document.

Hegland's AUTHOR program allows the producer of text to visualize and manipulate it in other than linear ways. The approach supports concept-focused writing, allowing the user to define as they write, and to then see the defined text in a Concept Map, while also exporting all defined text as an interactive Glossary.

The major contribution presented here is VISUAL-META. This is an approach for adding metadata to PDF documents, in a form that is equally readable to human and machine. Such data can then be interpreted and used to create properly formatted citations and afford more elaborate manipulations such as interactive graphs and charts. Hegland's insight is to add this information at the end of a PDF document and give it equal stature as the text of the document itself. In a sense, such a document "knows" itself and uses that knowledge to assist a reader. It is important that it enable both the human reader and software reader. VISUAL-META is entirely open for anyone to employ, with full specification being available on how VISUAL-META can contain citing information, structural/heading, glossary, endnotes, references and more, at the project website: <http://visual-meta.info>

Visual-Meta enables PDF readers, such as Hegland's READER, to augment the reader's ability to consume text in a variety of ways, such as the ability to copy text from the PDF and paste it in a Visual-Meta enabled

word processor, such as AUTHOR, where it will appear as a full citation, reducing the chance of errors. It also allows for novel views which basic PDF does not provide, both for a single document and for large volumes of documents.

These works are instances of what I think should be called *computational text*, by which I mean, text that lends itself to augmentation through computational tools. Engelbart referred to his research laboratory as the Augmentation Research Center^c to emphasize the capacity of computers to augment human capabilities and the concept seems applicable to augmenting the utility of text. One begins to visualize galaxies of content in a mixed-media universe and tools for production, discovery, and consumption. "But isn't that just the World Wide Web?" you might ask. Well, yes and no. The Web is indeed a remarkable and linked universe of wide-ranging content. Search engines and browsers aid in our ability to find and consume, render, and interact with that content. Hegland's tools add an exploitable self-contained self-awareness within some of the objects in this universe and increase their enduring referenceability.

It is exciting to learn the ACM Digital Library is exploring aspects of the VISUAL-META concept for incorporation into its operation. But then, one would expect ACM to look to cutting-edge ideas for the benefit of its members. □

^c https://en.wikipedia.org/wiki/Augmentation_Research_Center

Vinton G. Cerf is vice president and Chief Internet Evangelist at Google. He served as ACM president from 2012–2014.

Copyright held by author.

^a [https://en.wikipedia.org/wiki/NLS_\(computer_system\)](https://en.wikipedia.org/wiki/NLS_(computer_system))

^b <https://www.augmentedtext.info/>

The *Communications* website, <http://cacm.acm.org>, features more than a dozen bloggers in the BLOG@CACM community. In each issue of *Communications*, we'll publish selected posts or excerpts.



Follow us on Twitter at <http://twitter.com/blogCACM>

DOI:10.1145/3479972

<http://cacm.acm.org/blogs/blog-cacm>

New Life for Cordless Communication, Old Regrets for Software Projects

Andrei Sukhov and Igor Sorokin ponder the potential benefits of DECT to the Internet of Things, while Doug Meil considers how software engineers should reflect on their accomplishments.



Andrei Sukhov and Igor Sorokin

A Chance for Rebirth

<https://bit.ly/3lkg9lf>

July 1, 2021

The Internet of Things (IoT) is a new area of infocommunication technologies including not only home appliances with an IP address and Internet control, but also a variety of industrial technologies. These technologies need to be developed at all layers of the OSI model, but the need for security mechanisms should be emphasized.

At the physical level, wireless technologies (Wi-Fi, Bluetooth) have significant range limitations. The few tens of meters these standards allow are not enough. Alternative wireless technologies LPWAN (Low-power Wide-area Network) and NB-IoT (Narrow-band In-

ternet of Things) are rapidly developing and used in many areas. Their range is an order of magnitude higher, but with no generally accepted standard, they are regulated by private companies. Yet unresolved issues remain, so attempts are being made to develop new technologies to lay the foundation of the IoT.

Some older wireless technologies are well studied but in low demand, as interest in them has passed. We can try to give such technologies a second life in IoT, such as DECT (digital enhanced cordless telecommunication), which operates in the 1.9GHz band.

After the emergence of the GSM standard, interest in DECT dropped. An attempt to use DECT as a physical layer for TCP/IP was unsuccessful, due to low transmission rates. Emerging demand for IoT technologies gives DECT a chance for rebirth, as it can increase the communication range between devices up to several hundred meters and achieve a data transfer rate of several

hundred kilobits per second, sufficient for transmitting control and monitoring information in real time.

Note the advantage of DECT in comparison with existing standards for the IoT's physical layer. The frequency band in which it operates lacks many different devices, compared to the 2.4GHz band and the ISM (Industry, Science, Medicine) bands. In 2020, the EU adopted the DECT-2020-NR standards package to support IoT applications¹. Before that, in 2017, the DECT ULE (Ultra Low Energy) standard was adopted.

The advantages achieved when using DECT technology include:

1. Increased action at a distance of up to 600 meters.
2. Availability of ULE technology to extend service life.
3. Technologies to protect data transmission implemented at the physical layer of the OSI model (radio path level).
4. The ability to build a network infrastructure for mobile terminals without losing communication during transitions between base stations (handover).
5. A large number of standards approved by ETSI in 2020.

An IP over DECT data network should be a key component of a full-fledged family of IoT technologies. The ability to transfer data over the IP protocol, and to assign an IP address to a subscriber terminal, are the basic elements of the proposed concept. VoIP support has long been included in the DECT standard.

In some implementations of DECT phones, the subscriber device is an Android device that supports the TCP/IP protocol stack. The device is assigned an

IP address and can use all the capabilities of the IP protocol. There are many IoT MQTT (Message Queue Telemetry Transport²) implementations available for the Android platform, making testing easier. A DECT base station can be implemented based on a standard Linux server with a specialized PCI card.

In the cases described here, there is no need for high data transfer rates, and the speed at which voice transmission is organized is sufficient for IoT tasks.

The market presence of Android devices with DECT support makes it easy to assemble an experimental bench for tests and measurements. The authors anticipate the appearance of DECT subscriber devices based on minicomputers with a DECT module running Linux and Android operating systems. The emergence of such devices will mark the beginning of development as full-fledged technologies for IoT. The simplest smart plug devices based on DECT are already on the market, such as AVM's FRITZ! DECT technical product line.³

Our team plans to assemble a full-fledged DECT over IP stand, and to start developing and testing various IoT technologies. We are open to cooperating with other research groups on this.

References

1. ETSI. DECT-2020 New Radio (NR); Part 1: Overview; Release 1. European Telecommunications Standards Institute, Technical Specification (TS) 103 636-1, July 2020.
2. Boyd, B. et al. Building Real-time Mobile Solutions with MQTT and IBM MessageSight. *IBM Redbooks*, 2014.
3. <https://en.avm.de/products/fritzdect/>



Doug Meil
**Software Learning:
 The Art
 Of Design Regret**
<https://bit.ly/2TNkzFN>
 August 2, 2021

“What If?” is a question so fundamental to human learning that it has infused generations of science fiction enthusiasts with the possibilities of fixing our mistakes through time travel; 19th-century writer H.G. Wells' *The Time Machine* and “Star Trek”'s City on the Edge of Forever episode come to mind. Given time machines are not on the horizon, the best we can do is to look backward for insight and apply lessons forward. This is not as easy as it sounds, as there are two equally unhelpful poles: the first is to never to look to the past and question what could have been improved, and

the second is to persistently ruminate in the past. The goal is to live somewhere in the middle. How should software engineers try to classify their reflections?

Retrospectives have long been part of software engineering practice. It can be tempting to look at prior efforts and label every decision a “flaw” or “bug” if it doesn't agree with one's sensibilities. This is not constructive; understanding the context of decisions is critical, since no effort exists in a vacuum. Factors such as budget, resources (in quantity, quality, experience, and personality), technical options, and schedule all affect the decision-making process. Only by understanding design context can we differentiate between the preventable and the unavoidable, and understand what we could reasonably beat ourselves up about when we need to just let go.

Anachronistic Regret

This is where a framework or technology options did not exist at the time of a design decision, but regret is felt for not having those options anyway. While this can make for some interesting hypothetical discussions, such as the effects PCs could have had on the 1960s space race, it can also be taken too far, such as pondering how Abraham Lincoln could have revolutionized space travel as President if he only had rockets and computers during his administration. There was no decision possible, because there were no valid options at that time.

Actual Mistake Regret

It happens, where “it” can be everything from fat-fingers, code-horrors, or bear-traps, and Refactoring (Fowler). An example I lived through was a colleague enamored with the Inversion of Control and mocked objects pattern; there were unit tests, but of mocked objects instead of the codebase. When the codebase was deployed, it was a disaster, and I had to clean it up. The issue was not that Inversion of Control and mocked objects were not legitimate patterns; they were taken too far, as adherence to patterns became the goal, instead of the functionality of the overall software effort.

Decision Regret

This is the regret of the “road not taken.” In software efforts, there are design choices that seem ever-pitted against each other, such as natural vs. synthetic

keys for database design. Yet there are plenty of other cases where multiple reasonable options could exist for a situation that are not so doctrinally charged, each valid and “appropriate enough.” As long as each option was evaluated honestly and thoroughly in terms of strengths and weaknesses—and, ideally, documented—this is really the best we can expect any software engineer to do. Decision regret is an inevitable outcome of making decisions, and it is better to make progress and live with Decision Regret than to be paralyzed by it.

Unknown Consequences Regret

This is the case of “I wish I would have known that at the time,” where one experiences an unanticipated side effect or edge-case of a design, which often pops up at the worst possible time. These do not necessarily invalidate an overall design, but expose extra conditions that need to be addressed. The Java programming language is filled with these, particularly around memory management and garbage collection. Java has proven itself effective in a great many cases, but it also contains some surprises for designs that require operating under high memory load, where software solutions tend to work...until they don't. This necessitates diving into arcane Java Virtual Machine settings, and sometimes redesigning some software elements in response. In fairness to Java, every programming language and technology framework has sharp edges lurking somewhere, and finding those edges is the frequent consequence of doing interesting work.

Missed Opportunity Regret

Who hasn't exclaimed, “Why didn't I think of that?” Well, you didn't, and that's life. Again, the best one can do is to continually strive to expand one's knowledge horizons and look for opportunities to apply those lessons in the future. Fortune favors the brave—and the prepared. Iteration is equally important, as the more one practices, the better one can become at pattern recognition.

Andrei Sukhov is a professor and head of the Network Security Research and Study Group of HSE University, Moscow, Russia. **Igor Sorokin** is a postgraduate student in the Department of Computer Engineering of HSE University, Moscow, Russia. **Doug Meil** is a software architect at Ontada.

Flatter Chips

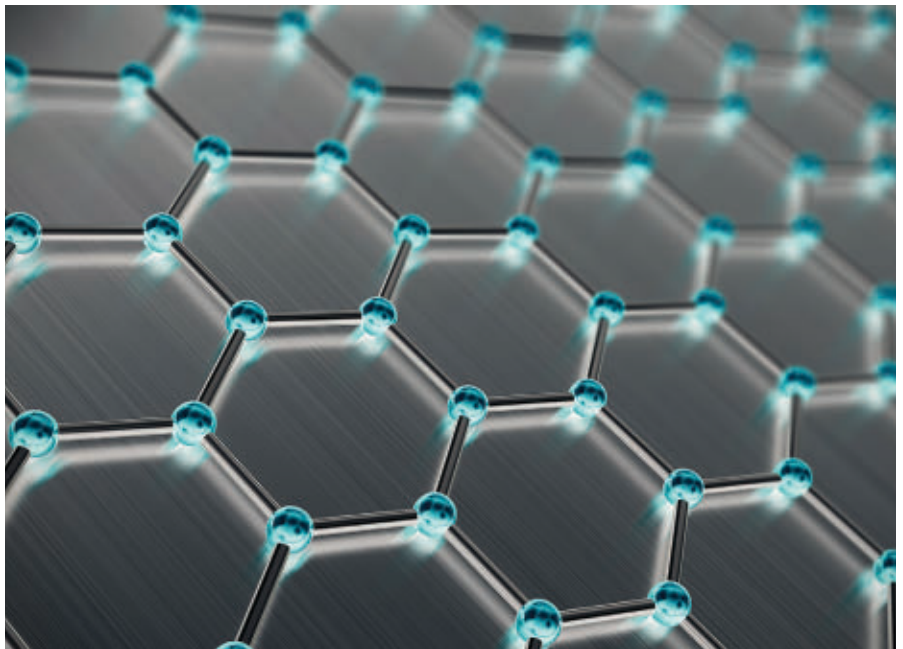
Two-dimensional materials—graphene and its cousins—could enable better integrated circuits.

MORE THAN 20 years ago, the historical rate of shrinking transistors to improve speed, density, power consumption, and cost became impossible to maintain. Even with slower physical scaling, however, electronics manufacturers steadily improved their products by exploiting new materials, new device and circuit designs, and faster communication between chips.

Many researchers believe there is a new opportunity to dramatically improve the transistors themselves by fashioning them out of the atomically thin sheets that naturally form in graphene and related layered materials. Incorporating a new material into highly optimized, state-of-the-art manufacturing will not be easy, although industry has done it before. It may start by grafting two-dimensional (2D) materials onto traditional chips to provide special capabilities, such as better interconnections or integrated optical devices. “Some form of 2D materials is going to be introduced integrally into the electronics eventually,” said Xiangfeng Duan of the University of California, Los Angeles (UCLA).

Graphene Dreams

Graphene—carbon atoms arranged in a chicken-wire-like sheet—has been



the archetypical 2D material since the mid-2000s, when U.K.-based physicists realized they could peel single graphene layers off a chunk of graphite using adhesive tape. (They received the Nobel Prize for their work in 2010.)

Graphene and several other 2D materials feature strong covalent bonds between atoms within the layers, which are in turn stuck together only weakly. This makes it easy to reproducibly create sheets of material that may be centimeters across but only one or a

few atoms in thickness, which are ideal for making short, fast field-effect transistors (FETs).

Moreover, because no bonds need be broken, the exposed surfaces are highly perfect, with few defects that disrupt electron motion within the layers. Electrons in graphene have a much higher mobility than they do in common semiconductors, which also translates into very fast transistors. Physicists continue to find exotic new effects in graphene, such as supercon-

ductivity that arises when two graphene sheets are slightly rotated relative to each other.

Graphene, however, has a fatal flaw, at least for large digital circuits: it is not a semiconductor. It is a semimetal, meaning that its conductivity never gets very small. Turning almost all of the billions of transistors in a chip *completely* off—with currents orders of magnitude smaller than when they are on—is essential to limit its power consumption.

Over the years, researchers have proposed various ways to transform graphene into a true semiconductor, such as preparing it as narrow nanoribbons. Manufacturing chips from such tiny filaments is hard, as it has proved for their cousins, carbon nanotubes.

“There’s lots of progress, but it’s an extraordinary challenge when you think about from the scaled integration point of view,” Duan said. “If you want six-nines yields [99.9999%], then it becomes very difficult to assemble these nanoscale-dimension materials, either nanoribbons or nanotubes.”

Still, academic researchers keep exploring new ways to manipulate graphene. Recently, Manoj Tripathi of the University of Sussex in the U.K. and his colleagues showed that tiny folds in a graphene sheet can change the local electronic structure, similarly to the doping changes that are used in traditional bipolar transistors. “We found that a single wrinkle is able to make a transistor,” Tripathi said, although he acknowledged it would be a long road to making useful circuits out of these tiny devices.

Beyond Digital Logic

“At the beginning, it was thought that graphene could be used in logic, but it’s not true. It’s a semimetal, so it’s always on,” said Amaia Zurutuza, chief scientific officer of Graphenea, a graphene supplier in San Sebastián, Spain.

Still, graphene sheets are already being used in other products.

“Most of the applications that we are pursuing (not directly, but in collaboration with customers) are combining graphene with silicon,” she said. For example, the Finnish company Emberion offers infrared and thermal imagers that use the semiconductor-like properties of graphene for light detection, placing the layers on top of

Graphene also could be useful for replacing or augmenting the wiring that connects transistors in a circuit, which also are formed after the devices are finished.

traditional silicon integrated circuits. The surface sensitivity of a thin graphene layer can also be exploited for detecting molecules, as well as light.

In the longer term, Zurutuza said, companies are exploring the possibility of coupling thin graphene devices with silicon photonics for telecommunications. “Sensors are considered lower-hanging fruit, because this requires back-end-of-the-line integration [after the sensitive and high-temperature processes that make the transistors], so the requirements for integration are not as stringent,” said Deji Akinwande of the University of Texas at Austin. Similarly, graphene-based memory might be stacked on chips late in the process.

Graphene could also be useful for replacing or augmenting the wiring that connects transistors in a circuit, which also are formed after the devices are finished. Ever-denser chips require ever-smaller wires, and often “graphene is more robust than copper,” Akinwande said. “It’s more reliable. It can handle more current. For a lot of the metrics that you care about for interconnects, the graphene-based material outperforms the existing copper, especially for scaled nodes.”

Beyond Graphene

“Scaled nodes” refers to the regular succession of integrated circuit technology generations, as prescribed by Moore’s law. In the early decades, these generations were named after the shortest gate electrode of FETs,

with other dimensions shrinking proportionally. That label lost its meaning around the turn of the century, however. “The whole idea of scaling is now some kind of ambiguous idea,” Akinwande said. “It becomes unclear whether this is just a marketing term.”

Unfortunately for researchers, since 2015 the semiconductor industry no longer publishes detailed specifications for future transistors. The revised document, the International Roadmap for Devices and Systems, “is more of a systems-level roadmap, so it doesn’t provide a lot of details for devices,” Akinwande said. “It’s kind of a free-for-all, at least in academia, how to push the boundary forward.”

Leading semiconductor manufacturers continued to drive dramatic progress in density and performance, but gate length stopped keeping pace with the label. One reason for this was the difficulty of shrinking the vertical dimensions of the transistor as it gets shorter, which allows the gate electrode to electrostatically control the channel to turn its conductivity on and off. Cutting-edge integrated circuits instead achieve this control using the FinFET device design.

Semimetallic graphene cannot exploit the improved control of short-channel effects in spite of its thinness, but there are numerous other layered materials that are intrinsically semiconducting. The most studied materials are the transition-metal dichalcogenides, abbreviated as TMDs or TMDCs, including molybdenum disulfide (MoS_2), tungsten diselenide (WSe_2), and indium diselenide (InSe_2) and many other materials. (Transition metals comprise the dozens of elements in the center of the periodic table, while chalcogens are the group VI elements sulfur, selenium, and tellurium.)

“Even five years ago, hardly any semiconductor industry company was investing in this, but now you see internal R&D is now emerging for trying to integrate these TMDs for scaled nodes,” said Akinwande. “They have very good transistor performance.” In contrast, he said, “As you make silicon thinner and thinner to make smaller transistors, the mobility falls off a cliff.”

A New Manufacturing Ecosystem

One important consideration in material choice, in addition to device per-

acm

Advertise with ACM!

Reach the innovators and thought leaders working at the cutting edge of computing and information technology through ACM's magazines, websites and newsletters.

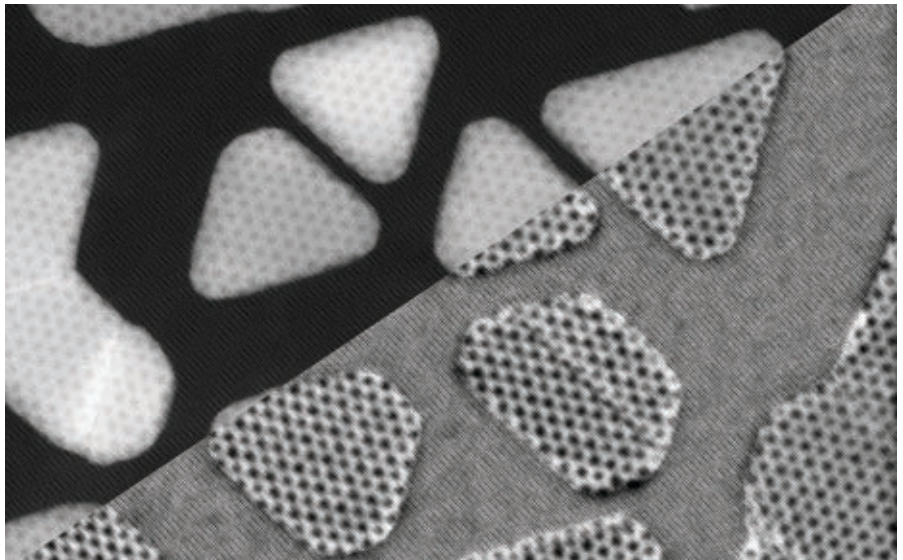


Request a media kit with specifications and pricing:

Ilia Rodriguez
+1 212-626-0686
acmm mediasales@acm.org

acm

media



"Islands" of gold atoms deposited on a layer of two-dimensional molybdenum sulfide, produced by Massachusetts Institute of Technology researchers using a new scanning transmission electron microscope.

formance, process complexity, and thermal stability, is toxicity, said Cedric Huyghebaert, program manager for exploratory materials and modules at the Interuniversity Microelectronics Centre (IMEC) in Belgium. As the technical leader of a new European project called 2D-EPL (Experimental Pilot Line), he noted that selenide processing uses very toxic chemicals. Because of the resulting environmental health and safety challenges if they were to be used on an industrial scale, the project is currently focusing on the sulfides MoS₂ and WS₂, in addition to graphene.

In most current explorations of 2D materials, "Typically they are grown on a template or on a catalyst metal, and then they are transferred" to the final wafer, Huyghebaert said. Because this process includes manual steps, "It brings a lot of variability, and this is not according to semiconductor standards."

"The 2D-EPL project is to try to set up an ecosystem" Huyghebaert said, to support and automate these processes, including materials suppliers, tool manufacturers, and device makers. "The product is only possible when you have the whole ecosystem."

By contrast, semiconductor manufacturers historically deposit materials as molecules onto partially processed wafers. "If you can grow directly these 2D materials onto an amorphous dielectric with the right quality, this will be the preferred option," Huyghebaert acknowledged, but so far it has not

been possible to get 2D materials of sufficiently high quality in this way, especially for graphene.

Moreover, he said, "The other reason why transferring could be quite interesting is that it would allow you to do heterostacking of these 2D materials," combining layers of different materials using techniques like those that have recently been developed to vertically package silicon devices. "Transferring very thin layers ... would be very powerful in the longer term for the semiconductor industry and nanotechnology."

For ultra-short transistors, Huyghebaert said, "The very appealing thing about TMDCs is that it would reopen the door for classical scaling for two to three generations." Still, he said, even if scaling is not successful, "There will be so many other opportunities that it will pay off in any case." **C**

Further Reading

Akinwande, D., Huyghebaert, C., Wang, CH., et al.

Graphene and two-dimensional materials for silicon technology. *Nature* 573, 507–518 (2019).
<https://doi.org/10.1038/s41586-019-1573-9>

Liu, Y., Duan, X., Shin, HJ., et al.

Promises and prospects of two-dimensional transistors. *Nature* 591, 43–53 (2021).
<https://doi.org/10.1038/s41586-021-03339-z>

Don Monroe is a science and technology writer based in Boston, MA, USA.

© 2021 ACM 0001-0782/21/10 \$15.00

Algorithmic Poverty

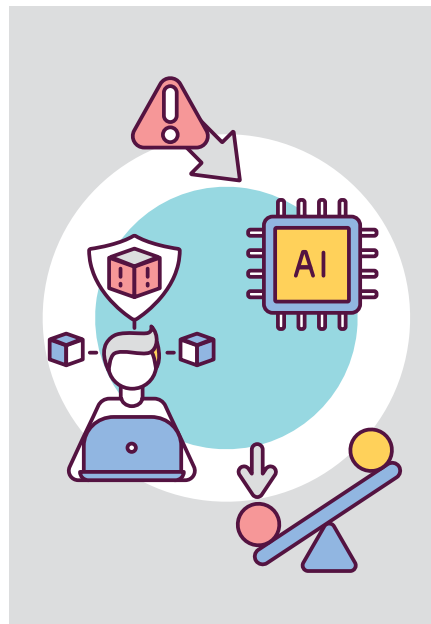
Algorithms can have a devastating impact on people's lives, especially if they already are struggling economically.

LIFE ISN'T FAIR" is perhaps one of the most frequently repeated philosophical statements passed down from generation to generation. In a world increasingly dominated by data, however, groups of people that have already been dealt an unfair hand may see themselves further disadvantaged through the use of algorithms to determine whether or not they qualify for employment, housing, or credit, among other basic needs for survival. In the past few years, more attention has been paid to algorithmic bias, but there is still debate about both what *can* be done to address the issue, as well as what *should* be done.

The use of an algorithm is not at issue; algorithms are essentially a set of instructions on how to complete a problem or task. Yet the lack of transparency surrounding the data and how it is weighed and used for decision making is a key concern, particularly when the algorithm's use may impact people in significant ways, often with no explanation as to why they have been deemed unqualified or unsuitable for a product, service, or opportunity.

"There are well-known cases of AI (artificial intelligence) and machine learning models institutionalizing preexisting bias," says Chris Bergh, CEO of DataKitchen, Inc., a DataOps consultancy. Bergh notes that in 2014, Amazon created an AI model that screened resumés based on a database of Amazon hires over 10 years. Because Amazon's workforce was predominantly male, the algorithm learned to favor men over women. "The algorithm penalized resumés with the word 'women' in references to institutions or activities (things like 'women's team captain')," Bergh says, noting "it took a preexisting bias and deployed it at scale." To Amazon's credit, once the issue was discovered, it retired the algorithm.

Perhaps the most serious com-



plaint about algorithms used to make decisions that impact financial determinations is that there is little transparency around the factors used to make those decisions, how the various elements are weighted, and what impact specific changes in behavior will have on improving outcomes. This is particularly devastating to those on the bottom rungs of the economic ladder; people seeking basic financial or medical assistance, housing, or employment may feel the impact of biased algorithms disproportionately, since being "rejected" for a product or service may actually be factored into the next algorithm they encounter. It is impossible to tell what the actual impact may be, because most firms keep their algorithms relatively opaque, because providing a fully transparent and open formula could allow users to inappropriately "game" the system and alter the algorithm's performance.

In 2019, Rep. Yvette Clarke (D-NY) introduced H.R.2231, the Algorithmic Accountability Act of 2019, which would direct the U.S. Federal Trade Commission to require entities that use, store,

or share personal information to conduct automated decision system impact assessments and data protection impact assessments. To date, no action has been taken on this bill.

Last year, the White House Office of Science and Technology Policy (OSTP) released a draft Guidance for Regulation of Artificial Intelligence Applications, which included 10 principles for agencies to consider when deciding whether and how to regulate AI. The draft noted the need for federal agencies that oversee AI applications in the private sector to consider "issues of fairness and non-discrimination with respect to outcomes and decisions produced by the AI application at issue, as well as whether the AI application at issue may reduce levels of unlawful, unfair, or otherwise unintended discrimination as compared to existing processes."

In the absence of laws or standards, companies may need to take the lead in assessing and modifying their algorithms to ensure the reduction or elimination of inherent or implicit biases. For example, Modern Hire, a provider of software used to streamline the hiring process via machine learning algorithms, has explicitly excluded certain elements from being used to assess a potential candidate during the hiring process, maintaining only those elements directly relevant to the position under consideration.

"The only things we score in the hiring process are things that the candidate consciously provides to us for use in the process," explains Eric Sydell, executive vice president of innovation for Modern Hire. "For example, we may take the audio of what they're saying [in an interview], and then we transcribe that into words. And then we score the words, and only the specific phrases and words that they actually verbalized."

Sydell says although Modern Hire has the capability to score candidates'

tone of voice, accent, and whether they sound enthusiastic, they do not score such attributes because those assessments could contain unconscious or conscious bias. Furthermore, there is not enough scientific evidence that such scores are effective indicators of new hire success. Says Sydell, “The science isn’t advanced enough at this point to score those things in a way that [eliminates biases].”

Another key strategy for helping to remove or reduce algorithmic bias is to ensure the group of people developing the model are from diverse backgrounds and have diverse perspectives of the world. “That’s how you can actually fix and balance models, and then you can make sure that you have different genders, different ethnicities, and different cultural perspectives, which are very, very important when you’re doing your model development,” says Seth Siegel, North American Leader of Artificial Intelligence Consulting for IT consulting firm Infosys. “You can never manage out all bias in a model, but what you can do is say, ‘okay, we have a huge gap in our training data model over here, so let’s go invest [into addressing that].’”

Still, relying on the traditional elements (credit scores, internal bank scores, past credit decision data, and the algorithms that tie this data together) used by landlords, banks, and other financial gatekeepers to assess an individual’s ability to pay rent, succeed at a job, or manage revolving debt, generally favors those who have managed to build up a successful track record of being assessed by those traditional institutions. Algorithms that assign more weight to the responsible use of traditional financial products and tools are likely to disproportionately impact people who are unbanked, disenfranchised, or otherwise outside of the financial mainstream, which often includes poorer people, minorities, or recent non-established immigrants.

“Today’s system may be fair for those inside it, but it is not inclusive,” says Naeem Siddiqi, senior advisor in the Risk Research and Quantitative Solutions division at business analytics software and services firm SAS. While Siddiqi has advocated for the use of alternative data to be incorporated into

“If you are waiting for an AI and machine learning model to tell you that, ‘Oh, you shouldn’t go do this’, it [won’t] happen.”

credit scoring models (such as historical utility payment data, rent payment data, or payments for things such as streaming services), he is not aware of any mainstream U.S. banks that do this at present.

It is inconceivable that large credit bureaus and the customers that utilize them will simply throw out the algorithms they current use, Siddiqi says. “[Although] building new credit risk models is not a huge undertaking, the bigger challenge is acquiring adequate alternative data, while following all the requisite privacy rules and regulations.”

That said, Infosys’ Siegel says some people simply don’t have great credit scores from a corporate risk perspective and as a result, they are unlikely to be provided access to the top tier of financial products and services. Still, “Companies that can figure out how to serve different parts of our society make money,” Siegel says. “There’s an incredible number of unbanked people in the U.S. Financial institutions that have offered similar banking products across [different] socioeconomic [levels], they perform better.”

Siegel says increasing pressure on these organizations to eliminate biases likely will lead some companies to use algorithms that do not rely on metrics or indicators that may include bias when they roll out a new product or service. But this approach still presents a massive challenge.

When designing and using algorithms, it is virtually impossible to weed out all sources of bias, because humans are the designers, approvers, and users of algorithms, and humans themselves have inherent biases, implicit and explicit, that are hard to ful-

ly eliminate. That’s why David Sullivan, a data scientist at data science and AI consulting startup Valkyrie, has taken an approach to managing algorithmic bias that flies in the face of conventional wisdom. Sullivan says algorithms are constructed to find relationships between historical trends, and it is the data and history being encoded that contains the prejudice.

“A counterintuitive, yet effective, way to address this bias in the data is to include protected classes in the data used to develop the algorithm, so that the scientist can control for that factor,” Sullivan explains. “The intention of including the data on the protected classes is to allow the model to encode what portion of the historical trend being modeled is based on those protected classes, and then exclude that relationship when making predictions on new data. This gives the model an ability to measure the historical impact of prejudice based on those protected classes, and explicitly avoid making predictions that rely on statistics affected by that bias.”

Sullivan adds, “It is only by being thoughtful and observant with our own history of prejudice that we can overcome it; this is as true with machine learning as it is with our own behavior.”

Indeed, the impetus is on the humans who need to take the necessary steps to assess and, if necessary, adjust their algorithms. “You have to actually take conscious action; don’t let models make all your decisions,” Siegel says. “If you are waiting for an AI and machine learning model to tell you that, ‘Oh, you shouldn’t go do this,’ it [won’t] happen.” **■**

Further Reading

Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms, Brookings Institution. May 22, 2019, <https://brook.gs/3rVY4ew>

O’Neil, C., *The Truth About Algorithms*, Royal Society for arts, manufactures and commerce, October 17, 2018, <https://www.youtube.com/watch?v=heQzqX35c9A>

Keith Kirkpatrick is Principal of 4K Research & Consulting, LLC, based in New York City.

A Switch in Time

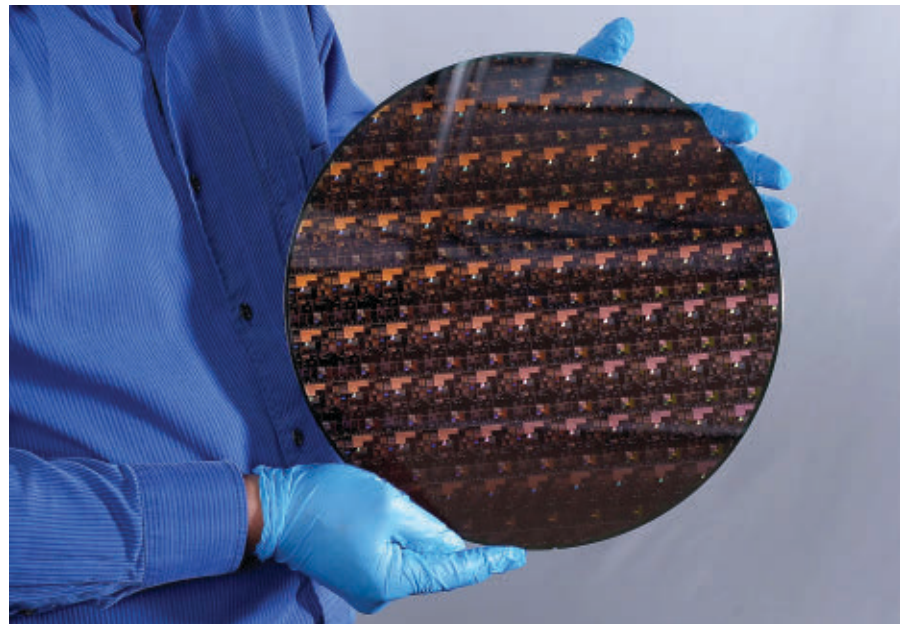
The introduction of the nanosheet transistor is just one step in the continuing attempt to stick to the spirit of Moore's Law.

SOUTH KOREAN CHIPMAKER Samsung Electronics aims to be first to adopt a new form of transistor that should allow Moore's Law to continue for another decade when it puts into production its 3nm semiconductor process toward the end of 2022.

It is just over a decade since the last major change to transistor structure went into production. The fin field-effect transistor (FinFET) emerged when the planar transistor structure that had served the industry well for several decades hit a physical limit. The problem lay in the relatively simple structure of the transistor's gate, an electrode laid over a thin channel between the source and drain that acts as an electrostatic valve. The gate's electric field, generated when a voltage is applied to it, controls whether electrons can pass through the channel, determining whether the transistor is switched on or not.

By the mid-2000s, chipmakers had succeeded in going beyond some expectations set by Moore's Law for the length of the gate. The 65nm node featured gates as short as 30nm that could switch quickly but suffered from high leakage. Charge carriers not only tunneled easily through the supposedly insulated gate, electric field lines generated from the drain were reaching the source region. That caused current to flow even when the transistor was supposed to be completely off. For several generations, gate length scaling stalled, to the point where chipmakers risked running out of space to place the conductive contacts needed to wire transistors to each other.

Beginning with the 22nm node, chipmakers switched to the FinFET. By lifting the transistor channel above the surface of the silicon, the gate electrode could be wrapped around three of its sides, instead of simply covering just the top surface, resulting in greater electrostatic control over electron



Last spring, IBM unveiled the first 2nm chip, built at its research facility in Albany, NY, USA.

flow. Now, even FinFETs are suffering issues similar to those of planar transistors a decade earlier. Surrounding the gate only on three sides still leaves an opportunity for some channel leakage to occur. The next step is to lift the channel above the silicon surface completely so the gate can wrap around the bottom as well.

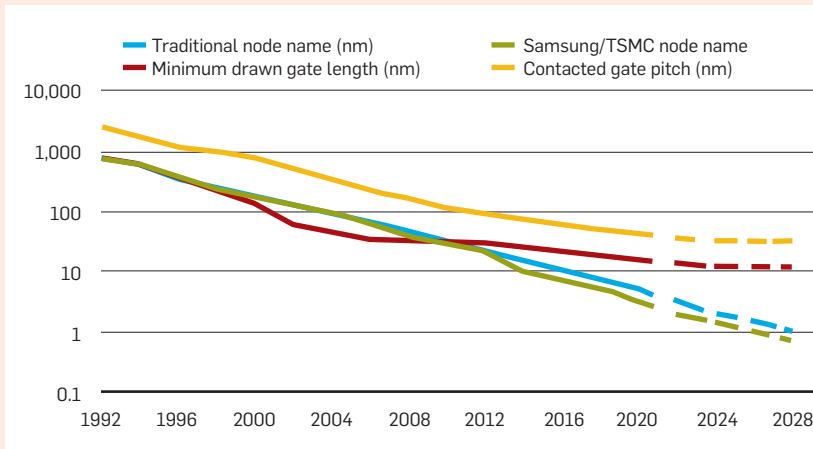
While there are a number of possible gate all-around (GAA) structures, manufacturers like Samsung favor the nanosheet design, a structure proposed by CEA-Leti and IBM 15 years ago. It involves some steps that are challenging, but has the advantage of being able to reuse many of the steps used to build FinFETs. The result is not just one enclosed channel, but several stacked on top of each other: an approach that further improves the control exercised by a wraparound gate. In place of the original silicon fin, a sandwich of multiple silicon and silicon-germanium layers is grown. Silicon germanium is used as a sacrificial layer because it provides an easy target for a chemical etch that can dis-

solve those layers away, to be replaced by the gate materials.

The horizontal form factor of the nanosheet provides an easier way to tune the size of the transistor. A major issue with the FinFET is that in most cases, a single fin in a transistor rarely provides enough current to be useful in a circuit. Multiple fins have to be used in parallel, so the effective width moves up in comparatively large steps. In his presentation at the International Solid State Circuits Conference in February, Samsung vice president of design enablement Taejoong Song said his team has taken advantage of the ability to draw nanosheets with different widths to create denser and more reliable memory cells than is possible with FinFETs.

A further boost will come in the form of energy efficiency. Chipmakers will take advantage of the improved gate control to reduce the supply voltage. As active power consumption is proportional to the square of the supply voltage, the savings to be made here can be substantial.

Divergence of key scaling factors from common node names up.



The International Roadmap for Devices and Systems (IRDS), an organization that has tracked semiconductor technology for more than two decades and which provides pathfinding data for chipmakers, expects the few manufacturers still able to make leading-edge silicon to have transitioned to nanosheet structures by the middle of this decade. But they are not all moving at the same time.

Expecting initial production from its competing process by the end of this year, the world's largest semiconductor foundry, Taiwan Semiconductor Manufacturing Company (TSMC), has opted to stick with FinFETs for one more generation, claiming it can still deliver 70% better density compared to the preceding N5 or 5nm process. The Taiwanese company will move to nanosheets for the N2 or 2nm process

that it hopes to debut by 2024.

Though the nanosheet brings benefits in terms of scaling, they will be far less dramatic than in the past. The IRDS estimates 12nm to be the limit for gate-length scaling for silicon-based transistors by 2030, a reduction of just 25% from what is achievable for 3nm nanosheet processes. There are also limits to how narrow they can become as well. Yet the IRDS still predicts an effective doubling in density according to Moore's Law to 2030 at least. The changes that allow scaling now have more to do with the way transistors are laid out and connected, rather than the dimensions of those devices.

For IRDS chairman Paolo Gargini, the changes the industry is making to achieve further scaling mark a return to what Gordon Moore said needed to take

place during his keynote at the IEEE International Electron Devices Meeting (IEDM) held more than 45 years ago. "If you go back to the 1975 presentation, he said the largest contribution to scaling would be from what he called 'circuit and system cleverness,' and that is what we will be doing in the upcoming decade," Gargini says. In today's terms, Moore's prediction could be restated as "transistors will be evolving into clever topological 3D structures," he adds.

That stronger focus on transistor layout and interconnection has been building for some time. It is the main reason why the names of process nodes have become increasingly disconnected from physical dimensions on-chip. Whereas in the 1990s, the node name generally reflected metal half-pitch or the gate length, the moniker *3nm* used by foundry suppliers Samsung and TSMC does not reflect any on-chip measurement. Even Intel's more conservative numbering of 5nm is still some way off from the actual gate length, which is at least three times longer.

Finding it difficult to reduce the spacing between parallel fins, chipmakers worked over the past decade to eliminate other sources of wasted space, such as how connections are made between transistors. Traditionally, the electrical connection to a gate would be placed to the side to avoid the risk of creating short-circuits with the source and drain connections. Intel found a chemical process that could reliably place the contact directly on top of the gate, making it possible to pack transistors

Milestones

2021 Knuth And Gödel Prizes Awarded

ACM's Special Interest Group on Algorithms and Computation Theory (ACM SIGACT) recently announced Moshe Vardi, of Rice University, will receive the 2021 Knuth Prize for outstanding contributions that apply mathematical logic to multiple fundamental areas of computer science.

ACM SIGACT also announced that the 2021 Gödel Prize was being awarded to five researchers: Andrei Bulatov of Simon Fraser

University; Martin E. Dyer of the University of Leeds; David Richerby of the University of Essex; Jin-Yi Cai of the University of Wisconsin, Madison; and Xi Chen of Columbia University, for their work on constraint satisfaction, a vital area of study within theoretical computer science.

Vardi's work has greatly increased understanding of myriad computational systems, and led to practical applications such as industrial hardware and

software verification. The major themes of his contributions are the use of automata theory and logics of programs to algorithmically prove correctness of system designs; the analysis of database issues using finite-model theory; characterizations of complexity classes such as P in terms of logical expressions; and the analysis of multi-agent systems such as distributed computation systems, via epistemic logic.

Bulatov, Dyer, Richerby, Cai,

and Chen were recognized for making important advances to the understanding of constraint satisfaction. Individually or in pairs, the researchers published papers on the classification of counting complexity of constraint satisfaction problems (CSPs) and proving an all-encompassing Complexity Dichotomy Theorem for counting CSP-type problems that are expressible as a partition function.

closer together without changing their internal dimensions. At the same time, chipmakers worked to reduce the number of parallel fins needed by making them taller, finding ways to reduce the risk of them collapsing during manufacture.

The industry now is looking to more radical changes in the layout of circuitry that surrounds the core transistors, further increasing the gap between the names given to process nodes and physical dimensions of the actual structures found on-chip.

Several years ago, originally as part of its proposal for N3 or 3nm-class processes, Belgian research institute Imec proposed burying power-supply lines under the transistor layer. Today, the power-supply lines interfere with logic routing, not least because they need to be relatively large so current pulses caused by high-frequency switching do not distort or break them.

Though burying the power rails can seem an obvious choice from the circuit designer's perspective, it is not an easy one for chipmakers to make. Benjamin Vincent, senior manager of the semiconductor process and integration at Lam Research subsidiary Coventor, says bringing metals into the production flow at that point "is something that the entire semiconductor industry has been avoiding for decades." The high-conductivity metals that will be needed can easily contaminate silicon surfaces, disrupting transistor formation.

By the end of the decade, the IRDS committee expects the industry to embrace not just buried power rails, but other ideas that pack transistors into a smaller area by exploiting the third dimension. CEA-Leti and Imec have recommended various methods for stacking transistors on top of each other. One leading candidate for the so-called 1.5nm process is Imec's CFET, which places the two complementary transistors used for most of today's logic in a vertical stack to achieve a near-50% saving in area.

There is a precedent for extensive vertical integration. Flash-memory suppliers demonstrated they can stack more than 100 memory cells vertically. Similar structures may beckon for logic transistors, though it will require another wave of manufacturing innovation to pull off.

"With stacked approaches, all the critical-dimension control requirements we had in previous technologies in the horizontal direction now move into the vertical direction," says Vincent. With this vertical 3D approach, no longer will the gate length be controlled by complex and expensive lithographic methods; instead, it will rely on accurate deposition of films to define the channel length.

Not surmounting these manufacturing challenges will likely bring Moore's Law to an earlier-than-expected finish. However, the IRDS committee and chipmakers see the renewed emphasis on topological "cleverness" in place of conceptually simpler area scaling as being the way to keep up with the decades-old law and pave the way to a 1nm process even if the gates, wires, and other structures on-chip turn out to be 10 times bigger than the claimed measurement. **Q**

Further Reading

Ye, P.D., Ernst, T., and Khare, M.
The last silicon transistor: Nanosheet devices could be the final evolutionary step for Moore's Law, *IEEE Spectrum*, Volume 56, Issue 8, August 2019.
<https://ieeexplore.ieee.org/abstract/document/8784120>

Song T. et al
A 3nm Gate-All-Around SRAM Featuring an Adaptive Dual-BL and an Adaptive Cell-Power Assist Circuit, *Proceedings of the 2021 International Solid State Circuits Conference*, pp338-340
<https://ieeexplore.ieee.org/document/9365988/>

Moroz, V. et al
DTCO Launches Moore's Law Over the Feature Scaling Wall, *Proceedings of the 66th IEEE International Electron Device Meeting (2020)*. pp41.1.1-41.1.4
<https://ieeexplore.ieee.org/document/9372010/>

Samavedam, S. B. et al
Future Logic Scaling: Towards Atomic Channels and Deconstructed Chips, *Proceedings of the 66th IEEE International Electron Device Meeting (2020)*. pp1.1.1-1.1.10
<https://ieeexplore.ieee.org/document/9372023>

International Roadmap for Devices and Systems – 2021 Update: More Moore
<https://irds.ieee.org>

Chris Edwards is a Surrey, U.K.-based writer who reports on electronics, IT, and synthetic biology.

© 2021 ACM 0001-0782/21/10 \$15.00

ACM Member News

WORKING AT THE BOUNDARY OF HCI AND AI



"I got into computer science by accident," says Meredith Ringel Morris, director and principal

scientist for People + AI Research (PAIR) at Google Research.

Before deciding on which college to attend, Morris toured the campus of Brown University in Providence, RI. The tour guide mentioned the boy in the movie *Toy Story* was named "Andy" after Andries "Andy" van Dam, a founder of Brown's computer science department and co-creator of one of the earliest hypertext systems in the 1960s.

Excited to learn that, "I signed up for Andy's class, which was my first exposure to computer programming," Morris recalls. "I loved it."

Morris went on to earn her undergraduate degree in computer science from Brown University. Both her master's and Ph.D. degrees were also in computer science, and both degrees were conferred by Stanford University.

On completing her doctorate, Morris joined the staff of Microsoft Research in Redmond, WA, where she remained until earlier this year, when she took the position with Google Research in Redmond.

Today, her research centers on Human-Computer Interaction (HCI), focusing on the design, development, and evaluation of collaborative, social, and accessible technologies. Morris works at the boundary of HCI and artificial intelligence to develop responsible AI-based technologies that can enhance the capabilities of all people.

"The PAIR team at Google Research focuses on new tools and techniques for human-AI interaction," Morris says.

She adds, "I hope that a human-centered approach to AI technology is recognized as being fundamental to the successful adoption of these complex socio-technical systems."

—John Delaney



DOI:10.1145/3481354

Michael A. Cusumano

Technology Strategy and Management

Section 230 and a Tragedy of the Commons

The dilemma of social media platforms.

AT THE CENTER of debate regarding regulation of social media and the Internet is Section 230 of the U.S. Communications Decency Act of 1996. This law grants immunity to online platforms from civil liabilities based on third-party content.²² It has fueled the growth of digital businesses since the mid-1990s and remains invaluable to the operations of social media platforms.⁶ However, Section 230 also makes it difficult to hold these companies accountable for misinformation or disinformation they pass on as digital intermediaries. Contrary to some interpretations, Section 230 has never prevented platforms from restricting content they deemed harmful and in violation of their terms of service. For example, several months before suspending the accounts of former President Donald Trump, Twitter and Facebook started to tag some of his posts as untrue or unreliable and Google YouTube began to edit some of his videos. Nevertheless, online plat-

forms have been reluctant to edit too much content, and most posts continue to spread without curation. The problem with false and dangerous content also seems not to have subsided with the presidential election: Social media is now the major source of anti-vaccine diatribes and other misleading health information.²¹

Given the law, social media platforms face a specific dilemma: If they edit too much content, then they become more akin to publishers rather than neutral platforms and that may

We must raise the consequences of spreading misinformation or disinformation.

invite strong legal challenges to their Section 230 protections. If they restrict too much content or ban too many users, then they diminish network effects and associated revenue streams. Fake news and conspiracy theories often go viral and have been better for business than real news, generating billions of dollars in advertisements.¹

The reluctance to edit content also has created what economists and others describe as a “moral hazard.” This phrase refers to a situation where an individual or organization can take risky actions because they do not have to bear the full consequences of taking those risks, such as when there is good insurance or weak government oversight.² In this case, social media platforms can pass on highly profitable falsehoods with relatively minor adverse consequences given the protections of Section 230 and (so far) manageable financial penalties for violating digital privacy rules or even antitrust regulations.¹⁴

Yet moral hazard may not be a strong enough term to describe what



could happen. As my coauthors and I have written elsewhere,^{5,7} another motivation for platform businesses to self-regulate more aggressively is the potential for a “tragedy of the commons.” This phrase refers to a situation where individuals or organizations narrowly pursue their own self-interest, as with moral hazard, but in the process deplete an essential common resource that enabled their prosperity to begin with.¹¹ Think of the native on Easter Island who cut down the last tree from a once-bountiful forest to make a fire—and then left everyone with an island that had no more trees. With online platforms, we can view the essential common resource as user trust in a relatively open Internet that has become a global foundation for digital commerce and information exchange. User trust, especially in dominant social platforms such as Facebook and Twitter, as well as in online marketplaces like Amazon and their product reviews, has been declining for years.^{8,15} Face-

book, YouTube, and TikTok are now claiming to be more transparent in how their algorithms work to relay information and detect false accounts aimed to manipulate readers for political and other purposes. However, most of these efforts seem to have been superficial or temporary.⁴

Of course, it is not unusual for politicians to manipulate the media for their own ends. In this regard, Donald Trump has been compared to former Senator Joseph McCarthy of Wisconsin (1908–1957).¹⁹ To get out his message about the threat of a Communist conspiracy at all levels of government and society, McCarthy had to rely on newspapers and magazines, radio and TV interviews controlled by a few established companies, and public Senate hearings.¹² By contrast, Trump relayed his message mainly through television (the Fox network) and social media platforms. With the latter medium, network effects can supercharge the flow of information as well as nonsense and dangerous falsehoods at

nearly 123,000 miles per second—the speed of Internet data transmission. With modern technology, Trump was able to communicate directly and at will with 88 million Twitter followers and another 60 million subscribers and readers on other social media sites.²⁰ Clearly, Trump exploited social media with a mastery that would have awed Joseph McCarthy, who only made it to the Senate.

After the insurrection attempt and Trump’s ongoing allegations of election fraud, Twitter and Facebook, followed by Google YouTube and Snapchat, all suspended his accounts. The platforms claimed Trump had violated terms of service that prohibit content encouraging violence or criminal acts. Amazon also stopped hosting the right-wing Parler social media app that some Trump followers had gravitated to as an alternative. However, these measures came after the Capitol violence and multiple deaths, and attracted criticism from both liberals and conservatives.¹⁶

Both Republicans and Democrats have advocated for the repeal of Section 230, for opposite reasons. While president, Trump argued the online platforms were already editing content from him and other right-wing sources and so they should no longer be protected from lawsuits charging them with discrimination.⁹ As a presidential candidate, Joseph Biden argued for the repeal of Section 230 because he felt online platforms should be held responsible for disseminating false or misleading content, which Section 230 prevents.¹³ So what can we do about the current dilemma facing social media?

First, the U.S. Justice Department or the U.S. Congress must amend Section 230 to reduce the blanket protections offered online platforms. Digital businesses still need some Section 230 protections to facilitate the open exchange of information, goods, and services. Yet we would all benefit from a revision of Section 230 that allows the public to hold online platforms accountable at least for advertisements and profits tied to the willful dissemination of false and dangerous information. We need government guardrails not only to protect the public but also to protect online platforms from their worst tendencies—the temptation to give in to harmful or destructive content that generates billions of dollars in sales and profits. Even Mark Zuckerberg has acknowledged the problem of regulating digital content is too big for Facebook and other social media platforms to solve by themselves.³

Second, the social media platforms must become more systematic and transparent in how they detect and curb false information that might endanger the public and individual lives.⁷ Suspending accounts that openly promote violence or dangerous misinformation is one measure they have taken already, but platform companies must act faster and more frequently. Facebook's recent use of an external oversight board, which upheld the ban on Trump's account—albeit temporarily—is a step in the right direction, but it was slow to engage and we do not know how often Facebook will call on its services.¹⁰ In addition, to sort through the vast amount of everyday Web traffic, social media platforms have employed

thousands of human editors to work along with computers running artificial intelligence and machine learning algorithms. Going forward, these companies will probably have to invest much more in both human editors and AI technology.

Third, we must acknowledge that technology, government regulation, or even more effective self-regulation by online platforms cannot by themselves fix the deep intellectual problems and political polarization that now plague American society, fueled by social media. For example, just after January 6, 2021, *The Washington Post* interviewed a man whose wife had been one of the police officers defending the Capitol building. He admitted his mother had been one of the rioters: “My mother has always been a conservative evangelical with extreme religious beliefs. My childhood was shaped by her profound distrust in science, public education, and vaccines. And yet my mom’s political activities basically tracked the mainstream of the Republican Party.”¹⁷ In another case, *The New York Times* described a former Navy Seal trained in counterintelligence who had come to the Capitol protests but remained outside. A registered Republican, we are told he completely “bought into the fabricated theory that the election was rigged by a shadowy cabal of liberal power brokers who had pushed the nation to the precipice of civil war. No one could persuade him otherwise.”¹⁸

We need to better understand why so many Americans are so susceptible to misinformation from social media and other sources. I once thought better education in science as well as history and ethics would help people think more clearly. Yet many well-educated Americans, including scores of politicians in our federal, state, and local governments, still claim the presidential election was rigged or oppose vaccines and masks. Education alone, apparently, is not the answer.

And so we risk a potential tragedy of the commons, where trust in online platforms declines to the point where very few people believe what they read on the Internet. Companies will always pursue their own self-interests, but we must raise the consequences of spreading misinformation or disinformation that can result in death, destruction,

and broad damage to society. We need both citizens and government actors to persuade platform company executives and boards of directors to recognize the danger of destroying trust in a technology that has provided them—and most of us—with so many benefits. ■

References

1. Aral, S., *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health—And How We Must Adapt*. Currency, New York, 2020.
2. Arrow, K. Uncertainty and the welfare economics of medical care. *The American Economic Review* 53, 5 (1963), 941–973.
3. BBC News. Mark Zuckerberg asks governments to help control internet content. (Mar. 30, 2019); <https://bbc.in/3CE7sIQ>
4. Boyd, A. TikTok, YouTube, and Facebook want to appear trustworthy. Don't be fooled. *The New York Times* (Aug. 8, 2021).
5. Cusumano, M., Gawer, A., and Yoffie, D. Can self-regulation save digital platforms? *Industrial & Corporate Change*. Special Issue on Regulating Platforms & Ecosystems, (2021).
6. Cusumano, M., Gawer, A., and Yoffie, D. *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. Harper Business, New York, 2019.
7. Cusumano, M., Gawer, A., and Yoffie, D. Social media companies should self-regulate. Now. *Harvard Business Review* (Jan. 15, 2021).
8. eMarketer. Facebook ranks last in digital trust among users. *eMarketer.com* (Sept. 24, 2020).
9. Edelman, G. On Section 230, it's Trump vs. Trump. *Wired* (Dec. 3, 2020).
10. Feiner, L., and Rodriguez, S. Facebook upholds Trump ban but will reassess decision over coming months. CNBC Tech Drivers (May 5, 2021); <https://cnb.cx/3AuGP71>
11. Hardin, G. The tragedy of the commons. *Science* 162 (1968), 1243–1248.
12. Hofstadter, R. *Anti-Intellectualism in American Life*. Vintage, New York, 1963.
13. Kelly, M. Joe Biden wants to revoke Section 230. *The Verge* (Jan. 17, 2020).
14. Kira, B., Sinha, V., and Srinivasan, S. Regulating digital ecosystems: Bridging the gap between competition policy and data protection. *Industrial & Corporate Change*, Special Issue on Regulating Platforms & Ecosystems, (2021).
15. Khan, L. Amazon's antitrust paradox. *The Yale Law Journal* 126, 3 (2017), 710–805.
16. Kirkpatrick, D., McIntire, M., and Triebert, C. Before the Capitol riot, calls for cash and talk of revolution. *The New York Times* (Jan. 16, 2021).
17. Marshall, D. My wife guarded the Capitol. My mom joined the horde surrounding it. *The Washington Post* (Jan. 23, 2021).
18. Philipps, D. From Navy SEAL to part of the angry mob outside the capitol. *The New York Times* (Jan. 26, 2021).
19. Remnick, D. What Donald Trump shares with Joseph McCarthy. *The New Yorker* (May 17, 2020).
20. Riley, K. and Stamm, S. How Twitter, Facebook shrank President Trump's social reach. *The Wall Street Journal* (Jan. 15, 2021).
21. Shephard, K. Miami private school says teachers who get corona virus vaccine aren't welcome, citing debunked information. *The Washington Post* (Apr. 27, 2021).
22. *The Wall Street Journal*. Section 230: The Law at the Center of the Big Tech Debate. (Nov. 18, 2020); <https://bit.ly/2VzSNxD>

Michael A. Cusumano (cusumano@mit.edu) is a professor and deputy dean at the MIT Sloan School of Management and coauthor of *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power* (2019).

I thank Alexander Eodice, Annabelle Gawer, Gary Genster, Mel Horwitch, John King, Nancy Nichols, Gilly Parker, Xiaohua Yang, and David Yoffie for their comments on prior drafts of this column.

Copyright held by author.

Broadening Participation Broadening Participation by Teaching Accessibility

Strategies for incorporating accessibility into computing education.

IN NOVEMBER 2012, *Communications* published a column by Vint Cerf titled “Why is Accessibility So Hard?”² In the column, Cerf describes the difficulty in building adaptable computer interfaces that can meet the needs of people with a variety of physical disabilities, blindness, deafness, and motor-related disabilities. With the multitude of computing platforms and applications that people interact with every day, reliably handling accessibility seems like a complex problem. Interestingly, a comment by one of the readers of the column, Bryan Garaventa, who is a blind developer, succinctly countered: “Accessibility isn’t hard. It takes discipline, knowledge, and comprehensive testing to get right, which are all part of the educational process.” Put it another way, accessibility is complex, but once it is learned, it is not that difficult to make accessible applications. In reality, not every accessibility problem has been solved and accessibility is an active research area. Nonetheless, much is known and practiced about incorporating accessibility into computing platforms and applications.

In this column, we will approach accessibility from a complementary perspective. Computing educators should be teaching accessibility—the theory and practice of designing and building accessible computing platforms and applications—in order to broaden participation in our field. This means bringing in more women, Black, Indigenous, and People of Color (BIPOC) people, disabled people, and members



of other minoritized groups into our field. It has already been well argued by many that broadening participation in the computing field yields tremendous benefits across the discipline.⁸ With that given, what is needed are explicit strategies to make this happen. One such strategy is teaching accessibility that has a twofold effect—direct and indirect. Directly, it is well known that topics in computing that relate to social good, like accessibility, attract women and BIPOC people to the field.^{1,5} Indirectly, if more computing platforms and applications were accessible (especially those that support our profession

such as IDEs), then more disabled people would be able to join our field, thereby helping make it more diverse.

Why Accessibility Is Important

According to the World Health Organization, there are approximately one billion people in the world who have a disability. Of this group, 285 million have a visual impairment not correctable by glasses or contact lenses and, of these, 39 million are blind. There are 466 million people with disabling hearing loss. According to the Christopher and Dana Reeve foundation, approximately 1 in 50 Americans have some sort of paralysis or motor-re-



STUDENT RESEARCH COMPETITION
Association for Computing Machinery
Advancing Computing as a Science & Profession
SPONSORED BY Microsoft

ACM Student Research Competition

Attention:
Undergraduate and Graduate Computing Students

The ACM Student Research Competition (SRC), sponsored by Microsoft, offers a unique forum for undergraduate and graduate students to present their original research before a panel of judges and attendees at well-known ACM-sponsored and co-sponsored conferences. The SRC is an internationally recognized venue enabling students to earn many tangible and intangible rewards from participating:

- **Awards:** cash prizes, medals, and ACM student memberships
- **Prestige:** Grand Finalists and their advisors are invited to the Annual ACM Awards Banquet
- **Visibility:** meet with researchers in their field of interest and make important connections
- **Experience:** sharpen communication, visual, organizational, and presentation skills

Learn more:
<https://src.acm.org>

lated disability such as spinal cord injury, multiple sclerosis, stroke, or cerebral palsy. These numbers extrapolate to about 150 million people worldwide. From an industry perspective, this represents a large number of customers who may have limited access to their platforms and applications if their accessibility is ignored. Interestingly, accessibility has direct benefits for people without disabilities because everyone has limitations in certain situations, such as when driving a car or in a noisy airport. Alternative ways for input and output provided by accessible applications are valuable in those situations. Ensuring computing platforms and applications are accessible is the answer.

From a human-rights perspective, people with disabilities should have access to computing platforms and applications that can benefit their lives. Indeed, some of these people should be an integral part of the workforce that creates those computing platforms and applications. The UN Conventions on the Rights of Persons with Disabilities, signed by more than 160 countries around the world, makes this case.⁷ Laws such as the Americans with Disabilities Act in the U.S. and the European Accessibility Act in the E.U. provide a legal basis for companies and governments to make products and services accessible.

For the reasons described here, many tech companies are eager to hire employees who have some expertise in accessibility as program managers and developers because they want their products to be accessible out of the box. No doubt, companies would like computing programs in universities and colleges to teach more about accessibility. The Teach Access organization has as its mission increasing the number of colleges and universities that teach accessibility topics.

Who Teaches Accessibility?

Accessibility topics appear naturally in a number of course settings including Web design/development courses, software engineering, and human-computer interaction courses. Indeed, any course that addresses human-facing hardware and software can have accessibility topics embedded in them. Approaches to designing human-facing software, including user-centered and

participatory design, should address users whose abilities are limited. The concepts of universal design and ability-based design should also be covered in such courses. Columnist Richard Lerner taught a course on data compression that introduced students to Grade II Braille—a compressed form of Braille—as part of the historical background of data compression. Grade II Braille uses fewer characters per word than uncompressed Braille allowing users to read more quickly with their fingers and use less paper. Disability and accessibility topics can be weaved into almost any course, even introductory courses. Paula Gabbert, a professor at Furman University, teaches an introduction to computing with accessibility as a theme throughout the course.⁴

Several years ago, Shinohara et al. surveyed computing faculty at U.S. institutions to ask whether or not they taught accessibility topics, and if not, why not.⁸ The survey was sent to 14,176 faculty members from 352 institutions, of which 1,857 from 318 institutions responded. Of those that responded, only 375 (20%) indicated they taught accessibility topics. We can probably conclude the vast majority of those who did not respond to the survey probably do not teach accessibility topics. This results in a very small number, perhaps 2.5% who teach any accessibility topics at all. The two most cited reasons that respondents did not teach accessibility were that it was not part of the core curriculum and they did not know enough about the topic to teach it. Almost half (46.6%) of the respondents agreed or strongly agreed that accessibility should be taught. The computing workforce needs graduates who know about accessibility, but even within Teach Access member schools, less than 3% of engineering and com-

Accessibility needs to be taught throughout the computing curriculum.

puting technology course descriptions reference accessibility.

How to Learn about Accessibility

The World Wide Web Consortium (W3C) works with member organizations across the globe to provide standards, guidelines, and resources to ensure the Web is accessible to everyone. In 2020, the W3C, in partnership with UNESCO IITE, launched “Introduction to Web Accessibility” (see <http://edx.org>) as a free MOOC on edX to teach professionals across the globe the fundamentals of Web accessibility and the benefits (for people with and without disabilities) when all Web applications are accessible. The W3C course is a great place to start learning the fundamentals about accessibility and why the solution is not to make specialized applications for users, but rather to apply principles such as universal design and ability-based design in order to make applications accessible to all.

In order to teach accessibility at scale, we need a multipronged approach to address both learning and teaching accessibility. As addressed in Kawas et al.,⁶ faculty and computing professionals need time and incentives to first learn accessibility themselves and then become comfortable with integrating it into their work (courses or applications) through techniques like what they call micro professional development. Individuals can continue to learn more about accessibility by partnering with organizations like Teach Access, AccessComputing, or W3C’s Web Accessibility Initiative (WAI).

► **Teach Access** is a collaboration between educational institutions, technology companies, and advocates for people with disabilities with the mission “to address the critical need to enhance students’ understanding of digital accessibility as they learn to design, develop, and build new technologies with the needs of people with disabilities in mind.”

► **AccessComputing** has the mission to increase the participation of people with disabilities in computing fields and provides a vast repository of resources for learning more.

► **W3C WAI** provides an international forum for people interested in Web accessibility.

Through learning more about ac-

cessibility, individuals can understand that both form and function play a part in developing meaning; if navigation and structure are set up only as visual elements, the meaning behind content disappears when the visual differences no longer appear. The popular operating systems from Apple, Microsoft, Google, and Linux that we develop on have designed their infrastructures to enable developers to take advantage of accessibility features such as screen readers, switch control, and speech control. These features enable the use of devices like speech output, refreshable braille displays, and physical switches. However, if the applications designed on top of those platforms are not coded properly, then accessibility is lost and we have form without functionality.

To get started, Teach Access has developed a free online tutorial (see <https://bit.ly/3xB6oIlg>) about accessibility, published resources for incorporating accessibility into curricula, and sponsored a grant program to incentivize faculty to incorporate accessibility into existing courses and share those modules with the community at large.

Call to Action

Accessibility must be taught throughout the computing curriculum and the call needs to come from each one of us. A bottoms-down strategy or top-up approach is not enough to affect the necessary change; we need stakeholders at all levels to make accessibility a priority.

1. **Computing departments:** (including CS departments, information schools, community colleges, boot camps, and high schools). Make accessibility a priority in the curriculum. Join Teach Access. Provide leadership and institution level support.

2. **Teachers:** Be a lifelong learner and make accessibility your next topic. Develop new modules about accessibility in your courses.

3. **Students:** Ask your department or program to include accessibility in the curriculum. Discuss accessibility in student ACM chapters, or form an accessibility student group. Find opportunities to learn about and promote accessibility. Organize an accessibility hackathon.

4. **Industry professionals:** Embrace

accessibility and make it a part of the corporate culture across your organization. Join Teach Access, build accessible products and services, include accessibility knowledge as a job requirement, be aggressive.

5. **Users and advocates (allies):** Help advocate for universal design and accessibility for all. Recognize that there is a broad spectrum of user needs, and one day, you or somebody you know will be the one who needs those accessibility “features.”

Conclusion

As Maya Angelou said: “I did then what I knew how to do. Now that I know better, I do better.” As computing professionals, we can take this quote to heart with regards to accessibility. As computer scientists, educators, students, practitioners, and designers, we may not have always known about accessibility or how to be inclusive of all, but now that we know better, we can do better. ■

References

1. Carrigan, C.M. Yearning to give back: Searching for social purpose in computer science and engineering. *Frontiers in Psychology* 8, 1178. (Jul. 2017); <https://bit.ly/3yI891d>
2. Cerf, V.G. Why is accessibility so hard? *Commun. ACM* 55, 11 (Nov. 2012), 7; <https://bit.ly/3U1y07>
3. Gabbert, P. Teaching accessibility in a CSO class. *Journal of Computing Sciences in Colleges* 35, 7 (Apr. 2020), 11–20.
4. Guzdial, M. et al. A statewide survey on computing education pathways and influences: factors in broadening participation in computing. In *Proceedings of the Ninth Annual International Conference on International Computing Education Research (ICER '12)*. Association for Computing Machinery, New York, NY, USA (2012), 143–150; <https://bit.ly/3ABtWiI>
5. Kawas, S., Vonessen, L., and Ko, A.J. Teaching Accessibility: A Design Exploration of Faculty Professional Development at Scale. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, (2019), 983–989; <https://bit.ly/3fTMDi>
6. Ladner, R. The impact of the United Nations convention on the rights of persons with disabilities. *Commun. ACM* 57, 3 (Mar. 2014), 30–32; <https://bit.ly/3yHHcdW>
7. Shinohara, K. et al. Who teaches accessibility? A survey of U.S. computing faculty. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 2018, 197–202; <https://bit.ly/3UHQI8>
8. Wulf, W. Diversity in engineering. *Leadership and Management in Engineering* 1.4 (2001): 31–35; <https://bit.ly/2VA51q1>

Kendra Walther (kwalth@usc.edu) is a senior lecturer in the Information Technology Program at University of Southern California, Los Angeles, CA, USA, and serves as co-lead of the Teach Access Student Task Force.

Richard E. Ladner (ladner@cs.washington.edu) is Professor Emeritus in the Paul G. Allen School of Computer Science and Engineering at the University of Washington, Seattle, WA, USA. He is the principal investigator for AccessComputing.

Copyright held by authors.

► Michael L. Best, Column Editor

Global Computing Remaining Connected Throughout Design

Applying the unique experiences of designing technologies for vulnerable communities.

A PPROXIMATELY 3.5% OF the world's total population are migrants: 272 million in 2019. This number has continually increased over the past 25 years, from 2.8% in 1995 (174 million), and 3.2% in 2005 (221 million).² Nearly two-thirds of these migrants seek economic opportunities such as better employment. While this strategy has benefited millions of internal and international migrants, booming markets and rapid urbanization have resulted in a constant demand for cheap labor, and in some instances, cases of forced labor and human trafficking.¹ Computing technology has begun to play an increasingly critical role in every step of the migration journey, from pre-departure to transit to integration or return to one's home country. It might also be leveraged to play a role in enhancing the rights of migrants. This column presents four cross-cutting challenges to co-designing technologies for and alongside migrant communities, drawing on the experience of developing Apprise, a mobile phone application to support vulnerable migrant workers to report exploitative work practices.

Understanding the Context: Perceptions, Problems, and Opportunities

Effective solutions cannot be generated by simply identifying surface-level symptoms of problems, but rather require a deep understanding



A migrant in Samut Sakhon, Thailand, uses the Apprise app.

of the multidimensional root causes (political, economic, and social) that enable these problems to continue. A participatory approach is necessary to break away from external notions of a community's needs and avoid the

reoccurring issue of seeing technology as an instant fix for issues affecting development. By prioritizing community participation before design ideation, teams can work together to identify these underlying factors, re-

sulting in more people-centric rather than tech-centric solutions.

As an initial step in our development of Apprise, we undertook a stakeholder analysis to identify key role players in each of the sectors that we work in across South East Asia: fishing, seafood processing, manufacturing, sex work, forced begging, domestic work. Across our four-year engagement, more than 1,500 stakeholders provided valuable insight to inform and shape the conceptualization, design, development and evaluation of Apprise, including vulnerable workers, survivors of trafficking, and frontline responders (FLRs—those with mandates in assessing working conditions for signs of labor exploitation and human trafficking such as NGOs, community-based organizations, government officials, labor inspectors, and IGOs).⁵ These interactions aimed to identify the issues that FLRs and migrant workers faced, and if/how they believed computing could support them to overcome these issues.

Throughout these interactions, our aim was to amplify the voices of migrant workers in precarious work situations through representative and transformative means of participation.⁷ Technological solutions created even in a perhaps well-intended vacuum and then imposed upon other populations have become notorious for failing with their tech-centric rather than people-centric approach. For this reason, we began our conversations asking stakeholders if there was a role for technology to help in addressing the issues they had identified. In each sector, we sought to understand from workers which group of stakeholders they feel most comfortable confiding in about work conditions. These stakeholders needed to have access to workers, a mandate to perform outreach in vulnerable communities, and most importantly be seen as a trustworthy by workers. In fishing and seafood processing, we work with government labor inspectors and NGOs; in manufacturing, we partner with private auditors within supply chains; in sex work, we collaborate with NGOs and sex-worker-led CBOs.

Continued Participation throughout Design Cycles

Further building upon this initial participation, it was critical to include

While privacy and security are important for all tech users, these factors become even more critical for vulnerable populations.

intended users throughout the design and evaluation phases of system development. Using a participatory and value-sensitive design approach, we identified key values of autonomy and privacy that were critical for migrant workers and frontline responders.⁴ These values informed the design of the system, the data collected, and the security considerations protecting user data. In the many cases where stakeholders presented competing perspectives, these values were used as a basis for negotiations and in some cases to adjudicate between design options.

Other design considerations we were able to identify included the importance of developing simple interfaces that prioritized learnability and the capability of operation in environments with intermittent connectivity. Considering aspects of autonomy, privacy, and trust, we designed Apprise to screen for indicators of labor exploitation by having a migrant worker respond to a series of yes/no-worded questions on a FLR's mobile phone. These questions are self-administered in an audio format with a set of headphones, using a combination of positively and negatively worded questions to ensure that any onlooker would not be able to interpret a worker's responses. These answers are recorded and uploaded to the FLR's account for post hoc analysis, enabling them to identify sector-specific practices of exploitation and provide a repository of case data for further investigation. Apprise's worker-centric and inclusive

design practices were recognized by the Worker Engagement Supported by Technology (WEST) community and highlighted in its white paper "Realizing the Benefits of Worker Reporting Digital Tools."⁶

Privacy, Security, Legal and Other Risks

While privacy and security are important for all tech users, these factors become even more critical for vulnerable populations and particular attention needs to be paid to the risks that collected data can create for individuals or groups of migrants. Risk assessment and risk mitigation should be an ongoing process throughout the whole data life cycle, minimizing data that is collected, stored, analyzed, and how long it is retained. Well-defined data governance roles and processes are also required to facilitate effective and appropriate data sharing and usage.³ These policies should be considered as part of the initial design, and in consultation with migrants (among other stakeholders). Consideration should be paid not only to data breaches and leaks, but also to the impact subpoenas for information from government or private parties would have on intended users. Worker engagement and other feedback platforms must pay particular attention to legal, financial, and reputational risks, specifically with respect to defamation cases.

In the evaluation of our initial pilot in fishing and sex work sectors in Thailand, there had been consensus amongst FLRs and migrant worker communities that the ability to capture photos and attach them to interview responses would increase the effectiveness of Apprise. We workshopped this suggestion at our next stakeholder consultation, discussing privacy concerns, and developed sharing strategies that would limit how images are shared within teams. There was one lone voice that spoke up during the workshop, warning of the increased risk that this functionality would bring to migrant worker communities should a data breach occur or a subpoena be received to share data. At the end of the workshop and based on the value dams and flows method that informed our design, we decided not to include this

new functionality, despite the improvements it would make to a FLRs ability to follow up on certain cases. This example is indicative of the need for higher levels of care in risk assessments when designing with and for migrant populations. These assessments should be continual and inform what data to collect, store, analyze, and retain.

Sustainability and Scalability of Digital Tools

Apprise was purposefully designed to be easily scaled to numerous geographies and scenarios to facilitate scalability. Making questionnaires available in more languages is a straightforward process involving the translation and verification of the questionnaire and new languages can be rolled out quickly upon request to enhance replicability in any part of the world. Using the ILO's Indicators of Forced Labor as a standard for developing preliminary interview questions, new lists can be updated and modified for sector-specific applications through consultation with stakeholders. Apprise was also designed as a complementary tool to be integrated

into the existing workflow of FLRs and to not include any additional resource burdens that would inhibit implementation. It does not require the construction of additional infrastructure to facilitate its usage, as it leverages an existing, widely available, relatively low-cost technology (the mobile phone). After initial adoption and refinement of best practices, it does not necessitate further resources that would lend it to be unsustainable in the longer term. This makes Apprise an example of a low-cost, potentially high-impact solution that is both sustainable and scalable.

Based on our collective learning experiences throughout the process of developing Apprise, we call for researchers, funders, and developers to take a broad perspective on needs assessment, stakeholder participation, risk assessment, and sustainability in conceptualizing and designing tech solutions for development. While these people-centric approaches may include a significantly longer lead-in time than more tech-centric approaches, this deep understanding and shaping of the project sets it up to be responsive to

the financial, cultural, technological, political, and environmental context that it is rooted within. If these factors are considered from very inception, perhaps we can avoid much of the disconnect between intention and actual impact that mark this space. □

References

1. Benach, et al. Migration and 'low-skilled' workers in destination countries. *PLoS Med.* 8, (June 2011), e1001043; doi: <https://bit.ly/2VNaf1B>
2. IOM, World Migration Report 2020. International Organization for Migration, Geneva (Nov. 2019); <https://bit.ly/3ySx4PX>
3. Stalla-Bourdillon, S. et al. Data protection by design: Building the foundations of trustworthy data sharing. *Data Policy* 2 (2020); doi: <https://bit.ly/3stQV7H>
4. Thinyane, H. and Bhat, K. Supporting the critical-agency of victims of human trafficking in Thailand. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland, May 2019.
5. Thinyane, H. et al. Use of New Technologies for Consistent and Proactive Screening of Vulnerable Populations. United Nations University, Institute in Macau, Macau, Jul. 2020; <https://bit.ly/3lSGCXp>
6. West Principles. White Paper: Realizing the Benefits of Worker Reporting Digital Tools. (Mar. 2019); <https://bit.ly/2VJLK5d>
7. White, S.C. Depoliticising development: The uses and abuses of participation. *Dev. Pract.* 6, 1 (Feb. 1996), 6–15; doi: <http://10.1080/0961452961000157564>

Hannah Thinyane (hannah@unu.edu) is Principal Research Fellow, United Nations University Institute in Macau, China.

Copyright held by author.

ICCCQ

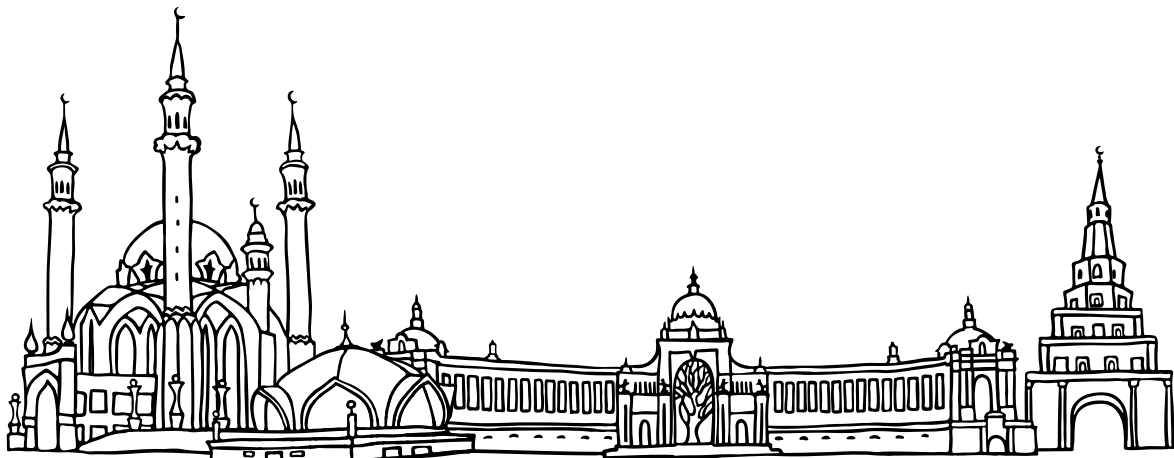
The Second International Conference
on Code Quality (23 Apr, online)

Static/Dynamic Analysis, Program Verification,
Bug Detection, and Software Maintenance

www.iccq.ru

CfP closes on 18 Dec

In cooperation with
ACM SIGPLAN and SIGSOFT
IEEE Computer Society





George V. Neville-Neil

DOI:10.1145/3481431

Article development led by [acmqueue](https://queue.acm.org)
queue.acm.org

Kode Vicious

Divide and Conquer

The use and limits of bisection.

Dear KV,

Many of our newer developers—those who have worked only with git—seem to find bugs in their code only by using git’s bisect command. This is troubling for a couple of reasons. The first is that often—once they find where the change occurred that caused the problem—they do not understand the cause, only that it happened between versions X and Y. The second is that they do not seem to understand the limits of debugging in this way, which, perhaps, is more a topic for you than for me to describe to you. Do you find this practice becoming more widespread and perhaps debilitating to good debugging?

Vivisected by Bisection

Dear Vivisected,

Nearly all new tools are both a blessing and a curse, as close readers of KV will know by now, and the ability to bisect a set of changes quickly is no different. It is quite definitely a blessing to have automation take over the tedious work of checking out a change, building the system, running a test, and seeing if the test fails, and then if it doesn’t fail in the right way, doing this all over again until the change that introduced the bug is found. That kind of work is something you want automated, and, therefore, in that case it is a blessing—a limited one, but a blessing nonetheless. I mean, it is not manna from heaven, is it?

What you are asking me to rant about (you are asking for a rant, right?) is how

such a tool can create lazy thinkers, and by extension, lazy engineers. Well, there are a few problems to talk about even before we get to whether having such automation leads to laziness.

Tools such as bisection are great if, and only if, you have a well-understood bug that occurs with 100% consistency so that the bisection can work. Bisection is of no use if you have a heisenbug, or something similarly subtle, that will fail only from time to time; and, while we do not want any bugs in our systems, we know these subtle bugs are the most difficult to fix and the ones that cause us—well, some of us—truly to think critically about what we are doing.

Timing bugs, bugs in distributed systems, and all the difficult problems we face in building increasingly complex software systems cannot yet be addressed by simple bisection. It is often the case that it would take longer to write a usable bisection test (the damnable thing you must write to get the bisection to tell you where the bad change was) for a complex problem than it would to analyze the problem while at the tip of the tree.

Another thing developers often fail to understand is the bug may not be related to any previous change; it might be right there in front of them, staring back, in orange on black. I have watched several developers who were absolutely convinced the bug was “somebody else’s problem” run and rerun bisections only to realize that the actual problem was in their latest, uncommitted change. It is unfair to laugh at people in

the middle of a debugging session, and, with KV, it is a risk to life and limb, but it is still damned tempting.

What bisection provides all developers is simply another tool to find bugs in their code. Sure, the bug must be easy to test for, likely cannot be in a distributed system, and cannot be a timing or a heisenbug, but it is still better than finding these simpler bugs by hand or writing your own script to do just what bisect is going to do.

Does this tool make us dumber? Probably not. What it does is perhaps help a less-seasoned developer to find bugs; however, if that developer wishes to learn, the tool does not prevent that, and that is why such a tool is a boon to some and a cushion to others.

KV

Q Related articles
on queue.acm.org

Kode Vicious

Debugging on Live Systems

<https://queue.acm.org/detail.cfm?id=2031677>

MongoDB’s JavaScript Fuzzer

Robert Guo

<https://queue.acm.org/detail.cfm?id=3059007>

Debugging Distributed Systems

Ivan Beschastnikh, Patty Wang,

Yuriy Brun, and Michael D. Ernst

<https://queue.acm.org/detail.cfm?id=2940294>

George V. Neville-Neil (kv@acm.org) is the proprietor of Neville-Neil Consulting and co-chair of the ACM *Queue* editorial board. He works on networking and operating systems code for fun and profit, teaches courses on various programming-related subjects, and encourages your comments, quips, and code snips pertaining to his *Communications* column.

Copyright held by author.

Viewpoint

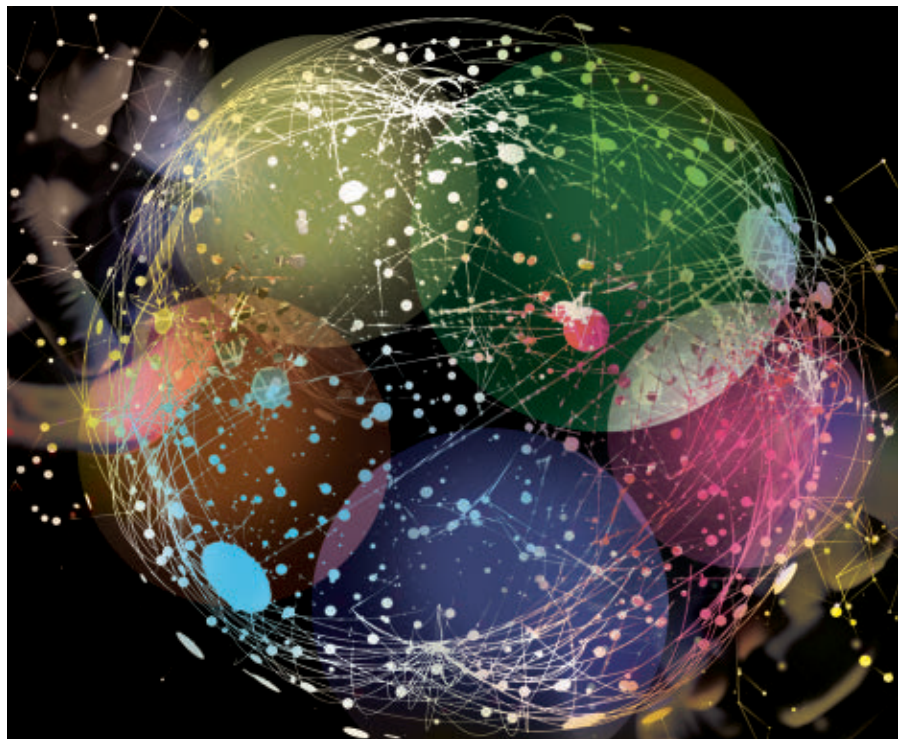
Competitive Compatibility: Let's Fix the Internet, Not the Tech Giants

Seeking to make Big Tech less central to the Internet.

TECH'S MARKET CONCENTRATION—summed up brilliantly by Tom Eastman, a New Zealand software developer, as the transformation of the Internet into “a group of five websites, each consisting of screenshots of text from the other four”—has aroused concern from regulators around the world.

In China tech giants have been explicitly co-opted an arm of the state. In Europe regulators hope to discipline the conduct of U.S.-based “Big Tech” firms by passing strict rules about privacy, copyright, and terrorist content and then slapping the companies with titanic fines when they fail to abide by them. At the same time, European leaders talk about cultivating “national champions”—monopolistically dominant firms with firm national allegiance to their local governments.

U.S. lawmakers are no more coherent: on the one hand, Congress recently held the most aggressive antitrust hearings since the era of Ronald Reagan, threatening to weaken the power of the giants by any means necessary. On the other hand, lawmakers on both sides of the aisle want to deputize Big Tech as part of law enforcement, charged with duties as varied as preventing human trafficking, policing copyright infringement, imposing neutrality on public discourse, blocking disinformation, and ending harassment and hate speech. If any of these



duties can be performed (and some of them are sheer wishful thinking), they can only be performed by the very largest of companies, monopolists who extract monopoly rents and use them to fund these auxiliary duties.

Tech has experienced waves of concentration before and resolved them with minimal state action. Instead, tech's giants were often felled by interoperability, which allows new market entrants to seize the “network effect”

advantages of incumbents to turn them to their own use. Without interoperability, AT&T ruled the nation. With interoperability, the ubiquity of the Bell System merely meant that anyone who could make an answering machine, radio bridge, or modem that could plug into an RJ-11 jack could sell into every house and business in America.

Everyone in the tech world claims to love interoperability—the technical ability to plug one product or service

into another product or service—but interoperability covers a lot of territory, and depending on what’s meant by interoperability, it can do a lot, a little, or nothing at all to protect users, innovation and fairness.

Let’s start with a taxonomy of interoperability.

Indifferent Interoperability

This is the most common form of interoperability. Company A makes a product and Company B makes a thing that works with that product, but does not talk to Company A about it. Company A does not know or care to know about Company B’s add-on.

You can find fishbowls full of USB chargers that fit your car-lighter receptacle at most gas stations for \$0.50–\$1.00. Your auto manufacturer does not care if you buy one of those \$0.50 chargers and use it with your phone. It is your car, it is your car-lighter, it is your business.

Cooperative Interoperability

Sometimes, companies are eager to have others create add-ons for their products and services. One of the easiest ways to do this is to adopt a standard.

Digital standards also allow for a high degree of interoperability: a phone vendor or car-maker who installs a Bluetooth chip in your device lets you connect any Bluetooth accessory with it—provided they take no steps to prevent that device from being connected.

This is where things get tricky: manufacturers and service providers who adopt digital standards can use computer programs to discriminate against accessories, even those that comply with the standard. This can be extremely beneficial to customers: you might get a Bluetooth “firewall” that warns you when you are connecting to a Bluetooth device that is known to have security defects, or that appears on a blacklist of malicious devices that siphon away your data and send it to identity thieves.

But as with all technological questions, the relevant question is not merely “What does this technology do?” It is “Who does this technology do it to and who does it do it for?”

The same tool that lets a manufac-

In the digital era, cooperative interoperability is always subject to corporate boundaries.

turer help you discriminate against Bluetooth accessories that harm your well-being allows the manufacturer to discriminate against devices that harm *its* well-being (say, a rival’s lower-cost headphones or keyboard) even if these accessories enhance *your* well-being.

In the digital era, cooperative interoperability is always subject to corporate boundaries. Even if a manufacturer is bound by law to adhere to a certain standard—say, to provide a certain electronic interface, or to allow access via a software interface like an API—those interfaces are still subject to limits that can be embodied in software.

What’s more, connected devices and services can adjust the degree of interoperability their digital interfaces permit from moment to moment, without notice or appeal, meaning the browser plugin^a or social media tool^b you rely on might just stop working.

Which brings us to ...

Competitive Compatibility

Sometimes an add-on comes along that connects to a product whose manufacturer is hostile to it: third-party inkjet ink, unauthorized iPhone apps, DVRs that record anything available through your cable package, and stores your recordings indefinitely.

Many products now have countermeasures to resist this kind of interoperability: checks to ensure you are not buying car parts from third parties,^c or fixing your own tractor.^d

When a manufacturer builds a new product that plugs into an existing

one despite the latter’s hostility, that is called “competitive compatibility”^e and it has been around for about as long as the tech industry itself, from the mainframe days^f to the PC revolution^g to the operating systems wars^h to the browser wars.ⁱ

All three forms of interoperability share some characteristics: in each case, technologists devise a means by which two or more products or services can extend one another’s functionality, read one another’s files, or otherwise provide benefit to the users of one or both services.

The difference between these forms of interoperability is in the type of technical work necessary to accomplish them.

Firms that create APIs or other interfaces to explicitly invite third-party add-ons contemplate both their users’ and employers’ priorities and try to strike a balance between them, crafting a means whereby their inventions can be improved or adapted by others without foregoing unacceptable future revenues from making such improvements on their own.

Firms that participate in standards-setting make a similar calculus but arrive at a different equilibrium. A multistakeholder format means that if you try to standardize, say, the costs of your products (in the hopes of getting others to shoulder them), while maintaining as proprietary the sources of your profits, you will have to convince other participants (including your commercial rivals) that this is a fair arrangement. Standards Development Organizations describe these compromises as a major feature of standardization itself: rivals check one another’s most greedy impulses and arrive at a fair middle ground that does not unduly advantage any one firm (of course, in highly concentrated markets, large firms can collude to create standards that advantage them at the expense of potentially disruptive new market entrants).

These “cooperative interoperability” efforts can give rise to follow-on, “indifferent interoperability” mo-

a See <https://bit.ly/3IVVPai>

b See <https://bit.ly/3AB99EQ>

c See <https://bit.ly/3IRnWax>

d See <https://bit.ly/3scBGNZ>

e See <https://bit.ly/3CEE1X7>

f See <https://bit.ly/2XekFrm>

g See <https://bit.ly/3xM6beN>

h See <https://bit.ly/3FYllaK>

i See <https://bit.ly/3yHKwZ>



ACM Transactions on Evolutionary Learning and Optimization (TELO)

ACM Transactions on Evolutionary Learning and Optimization (TELO) publishes high-quality, original papers in all areas of evolutionary computation and related areas such as population-based methods, Bayesian optimization, or swarm intelligence. We welcome papers that make solid contributions to theory, method and applications. Relevant domains include continuous, combinatorial or multi-objective optimization.



For further information and to submit your manuscript, visit telo.acm.org

ments. These occur when new products and services leverage deliberately interoperable technologies to do things that are orthogonal the considerations that went into the original. Think of the USB charger that plugs into a car's lighter receptacle: the firms that standardized the receptacle in the first place worked carefully to ensure none of their cars would be at a competitive disadvantage when it came to attracting drivers who smoked; they gave careful consideration to production, maintenance, and safety; but they did not even consider a distant future in which a universal power-cable would emerge to charge lithium-ion cells in commodity consumer electronics.

As tobacco smoking declined and device-charging grew, automakers gave more consideration to this new use case, and even encouraged it, turning indifferent interoperability into cooperative compatibility after the fact.

Unlike cooperative interoperators or indifferent interoperators, technologists engaged in competitive compatibility have an adversarial relationship with those who came before them. To defeat the anti-tampering chip in a single-use print-cartridge, or field a scraper that exports user-data from a giant's walled garden, or make a third-party office suite that seamlessly reads and writes an incumbent's spreadsheets, word processor documents and presentations, a technologist must defeat obfuscation, encryption, intrusion detection, and other countermeasures meant to thwart them.

The indifferent interoperator faces challenges that the cooperative interoperator does not. The cooperative interoperator can put in a request for an API extension, or argue in a standards committee for the inclusion of a feature they need. The indifferent operator has no leverage over the product's vendor(s), and has to work within the constraints of the product as it exists in the field.

Technologists who engage in competitive compatibility, however, are actively working at cross-purposes to those who came before them. They are playing a game of cat-and-mouse, relying on exploiting defects, or camouflaging their tools as normal user activities, and they must contend with

the possibility that the result of their efforts will be revisions to the original product or service explicitly designed to break their add-ons (indifferent operators sometimes see their work undone by these updates, but only as an incidental effect and not out of any animus to them).

Competitive compatibility can also collapse into cooperative compatibility. Sometimes dominant companies surrender and agree to cooperate: today's office file formats are standardized under ISO, the proprietary HTML extensions of the browser wars have been discarded or integrated into W3C standards, and so on.

There is a reason that compatibility tends to win out over the long run—it is the default state of the world—the sock company does not get to specify your shoes and the dairy does not get to dictate which cereal you pour milk over.

But as technology markets have grown more concentrated^j and less competitive, what was once business-as-usual has become almost unthinkable, not to mention legally dangerous, thanks to abuses of cybersecurity law,^k copyright law,^l and patent law.^m

Taking competitive compatibility off the table breaks the tech cycle: a new company enters the market, rudely shoulders aside its rivals, grows to dominance, and is dethroned in turn by a new upstart. Instead, today's tech giants show every sign of establishing a permanent, dominant position over the Internet.

“Punishing” Big Tech by Granting It Perpetual Dominance

As states grapple with the worst aspects of the Internet—harassment, identity theft, authoritarian and racist organizing, disinformation—there is a real temptation to “solve” these problems by making Big Tech companies legally responsible for their users' conduct. This is a cure that is worse than the disease: the big platforms cannot subject every user's every post to human review, so they use filters, with catastrophic results.ⁿ At the same

^j See <https://bit.ly/3scslWt>

^k See <https://bit.ly/3yDaMBr>

^l See <https://bit.ly/2XntxLw>

^m See <https://bit.ly/3iEvECP>

ⁿ See <https://bit.ly/2VPgO3c>

The biggest Internet companies need more legal limits on their use and handling of personal data.

time, these filters are so expensive to operate that they make it impossible for would-be competitors to enter the market. YouTube has its \$100 million Content ID copyright filter now, but if it had been forced to find an extra \$100,000,000 to get started in 2005, it would have died a-borning.

But assigning these expensive, state-like duties to tech companies also has the perverse effect of making it much harder to spark competition^o through careful regulation or break-ups. Once we decide that providing a forum for online activity is something that only giant companies with enough money to pay for filters can do, we also commit to keeping the big companies big enough to perform those duties.

Interoperability to the Rescue?

It's possible to create regulation that enhances competition. For example, we could introduce laws that force companies to open their back-ends^p and oversee the companies to ensure they are not sneakily limiting their rivals behind the scenes. This is already a feature of good telecommunications laws,^q and there is a lot to like about it.

But a mandate to let users take their data from one company to another—or to send messages from one service to another—should be the opener, not the end-game. Any kind of interoperability mandate has the risk of becoming the ceiling on innovation, not the floor.

Fix the Internet, Not the Tech Companies

The problems of Big Tech are undeni-

able: using the dominant services can be terrible, and now that they have broken the cycle of dominance and dethroning, the Big Tech companies have fortified their summits such that others dare not besiege them.^r

The biggest Internet companies need more legal limits on their use and handling of personal data. That's why we need a national privacy law, with a "private right of action" so that users can bring suit if they are victimized by surveillant companies. But laws that require filtering and monitoring user content make the Internet worse: more hostile to new market entrants (who cannot afford the costs of compliance) and worse for Internet users' technological self-determination.

If we are worried that shadowy influence brokers are using Facebook to launch sneaky persuasion campaigns,^s we can either force Facebook to make it more difficult for *anyone* to access your data without Facebook's explicit approval (this assumes that you trust Facebook to be the guardian of your best interests)—or we can bar Facebook from using technical and legal countermeasures^t to shut out new companies, co-ops, and projects that offer to let you talk to your Facebook friends without using Facebook's tools, so you can configure your access to minimize Facebook's surveillance and maximize your own freedom. That would mean reforming the Computer Fraud and Abuse Act to clarify that it cannot be used to make Terms of Service violations into civil or criminal offenses; reforming the Digital Millennium Copyright Act to clarify that defeating a technical protection measure is not an offense if doing so does not result in a copyright infringement; comprehensively narrowing software patents to allow for interoperable reimplementations; amending copyright to dispel any doubt as to whether reimplementing an API is a copyright infringement; and limiting the anticompetitive use of other statutes including those relating to trade secrecy, nondisclosure, and noncompete.

The second way is the better way.

Instead of enshrining Google, Facebook, Amazon, Apple, and Microsoft as the Internet's permanent overlords and then striving to make them as benign as possible, we can fix the Internet by making Big Tech less central to its future.

It's possible that people will connect tools to their Big Tech accounts that do ill-advised things they come to regret. That is kind of the point, really. After all, people can plug weird things^u into their car's lighter receptacles, but the world is a better place when *you* get to decide how to use that useful, versatile ANSI/SAE J56-compliant plug—not GM or Toyota.

Corporations Make Terrible Governments

AT&T was very nearly broken up in 1956. The monopolistic conduct that had enraged rural Americans and would-be telecoms rivals reached such an undeniable nadir that the DoJ finally moved to break up Ma Bell. Only one thing stood in the way: the Pentagon. AT&T had been deputized to perform so many state-like duties during its decades of monopolistic operations that it had acquired powerful stakeholders in the U.S. government—it had its own army! The Pentagon told the DoJ that it could not successfully occupy Korea an intact AT&T: it needed its Death Star to be a fully operational battle-station.

AT&T got a stay of execution, and instead was slapped with restrictions on its conduct that it skirted, violated, and flouted for the next three decades, until, finally, it was broken up in 1982. That 26-year reprieve was the direct result of "fixing" AT&T by trying to co-opt it to serve the state, rather than using the power of the state to weaken it.

Government derive their power from the consent of the governed. Their legitimacy comes from their accountability. Companies have shareholders, not citizens. Businesses are not governments, and they have no businesses governing us. □

^u See <https://bit.ly/3fZ2FYD>

Cory Doctorow (cory@eff.org) is Visiting Professor of Computer Science at the Open University, U.K.

Copyright held by held author.

^o See <https://econ.st/3yIGT2J>

^p See <https://bit.ly/3xD1DHD>

^q See <https://bit.ly/3CIUOYN>

^r See <https://bit.ly/2VJmRHe>

^s See <https://bit.ly/2Xj9C05>

^t See <https://bit.ly/3scYU9n>

Viewpoint

AI Futures: Fact and Fantasy

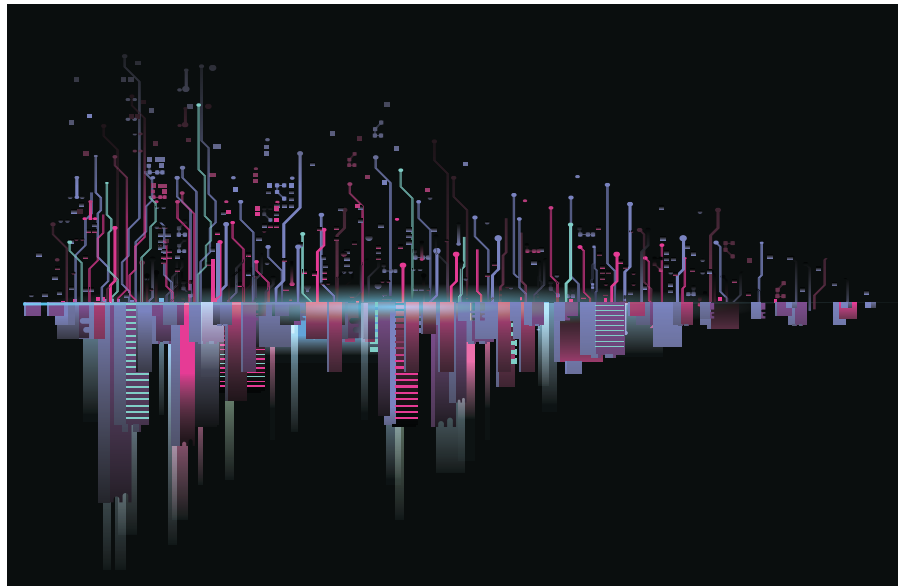
Three books offer varied perspectives on the ascendancy of artificial intelligence.

“ALPHAZERO CRUSHES CHESS!” scream the headlines^a as the AlphaZero algorithm developed by Google and DeepMind took just four hours of playing against itself (with no human help) to defeat the reigning World Computer Champion Stockfish by 28 wins to 0 in a 100-game match. Only four hours to recreate the chess knowledge of one and a half millennium of human creativity! This followed the announcement just weeks earlier that their program AlphaGoZero had, starting from scratch, with no human inputs at all, comprehensively beaten the previous version AlphaGo, which in turn had spectacularly beaten one of the world’s top Go players, Lee Seedol, 4-1 in a match in Seoul, Korea, in March 2016.

Interest in AI has reached fever pitch in the popular imagination—its opportunities and its threats. The time is ripe for books on AI and what it holds for our future such as *Life 3.0: Being Human in the Age of Artificial Intelligence* by Max Tegmark, *Android Dreams* by Toby Walsh, and *Artificial Intelligence* by Melanie Mitchell.^{6,8,9} All three agree on the boundless possibilities of AI but there are also stark differences.

First, their styles reflect perhaps the personalities of their authors—on one side is Max Tegmark, a professor of physics at MIT who communicates with high-flying folksy flamboyance.

^a <https://bit.ly/3yhaekI>



His book is hardly about AI at all, save one chapter where he quickly compresses how “matter turns intelligent.” For the rest, it ranges over vast magnitudes of space and time as befits a cosmologist: 10,000 years in one chapter, and if that is not enough, a billion years and “cosmic settlements” and “cosmic hierarchies” in the next.

On the other side are the books by Walsh and Mitchell who are more staid academics with feet firmly on the ground. Walsh and Mitchell are computer scientists who have worked in AI for a large part of their professional lives. Part 1 of Walsh’s book gives a survey of how AI developed from the seminal paper of Alan Turing, through the early days of GOFAI (“Good Old

Fashioned AI”) to modern statistical machine learning and Deep Learning. This is a fairly accurate evolution of the discipline and the different “tribes” in it, a compressed version of the account in Pedro Domingo’s *Master Algorithm*³ from a few years back. Part 2 and much of Mitchell’s book is a panoramic survey of the present state of the art in AI in areas such as automated reasoning, robotics, computer vision, natural language processing, and games. Both discuss the AlphaGo program—its strengths and limitations. Walsh voices some skepticism about how general the technique is, conjecturing that it “would take a significant human effort to get AlphaGo to play even a game like chess well.” The perils of forecast-

ing—just a year later AlphaZero used the same principles to crush chess with no human input. Subsequently, AI has beaten humans at Poker and defeated a pair of professional Starcraft II players. While Waymo has opened limited taxi services, the excitement about autonomous driving has recently cooled off somewhat and full services are still a way off.

As to the future of AI, all three agree that in principle, *superintelligence* is possible: that machines could, in principle, become more intelligent than humans, as indeed Turing contemplated in his original paper from 1950. For Tegmark, there are no physical laws that are violated and for Walsh and Mitchell, there are no computational principles that preclude it.

When is this likely to happen? Here the books could not be more different. For Walsh and Mitchell and the Stanford 100-year study of AI,^b this is today only a very distant possibility, several decades away if not more. Tegmark, on the other hand, seems to suggest this is just around the corner, and moreover that a sizeable number of AI researchers also think so. A poll conducted by Müller and Bostrom from the Future of Humanity Institute (FHI) is often cited as proof of this but subsequent polls that target a more informed group of researchers—namely those who had “made significant and sustained contribution to the field over several decades” came to very different conclusions. Another even more recent poll by the FHI group, this time targeting AI experts⁴ also came to somewhat more nuanced conclusions.

So, as AI advances rapidly, what are the future risks? Here again, they agree on a few things. All three are seized of the dangers of autonomous weapons and have devoted a lot of effort to lobbying AI researchers to sign a declaration against such weapons. All are cognizant of the threat of automation to jobs, though Tegmark mentions it only in passing in one section.

But for the most part, they are on totally different planes. As befitting a cosmologist, Tegmark is again thinking in grand terms: *Life 3.0* is life that can (more rapidly) change both its software and hardware. This is the stuff of movies like *The Matrix* or *Space Odyssey*

Interest in AI has reached fever pitch in the popular imagination.

and here Tegmark displays his considerable talents at fantasizing: a whole chapter is devoted to various types of *Matrix*-like future scenarios, some featuring benign AIs, others malignant. *Life 3.0* starts with a parable of a future with a HAL-like computer taking over the world. Perhaps a future career awaits Tegmark in the sci-fi movie industry. Tegmark also possesses great fund-raising talent—he has founded his own Future of Life Institute (FLI) devoted to these questions and secured a donation of \$10 million from Elon Musk who also likes to indulge in such speculations. There is an entire chapter in the book about the drama surrounding an event organized to announce the institute and the grant. One can see the need for hyperbole in such projects, but it borders on irresponsible to claim, as Tegmark has done, that AI is a more imminent existential threat than climate change: while there are precise projections of time frames from climate science, the former are purely speculative.

Noted roboticist Rodney Brooks has warned of the “seven deadly sins of AI predictions.”² In particular, he issues a warning about imagined future technology: “If it is far enough away from the technology we have and understand today, then we do not know its limitations. And if it becomes indistinguishable from magic, anything one says about it is no longer falsifiable.”

Life 3.0 is guilty of several of these deadly sins.

Walsh and Mitchell have concerns with AI risks that are of a totally different sort. They are sceptics about superintelligence and outline their own arguments why it may never ever be possible. They are not worried about superintelligent machines but rather super stupid machines with their bugs and failures and how we are reposing

faith in them. They are worried about systematic biases in AI systems and consequences for fairness when they are entrusted with decision making responsibilities. And they are concerned about the consequences of automation on jobs and the economy.

The AI community has started taking the risks of AI seriously and there are whole themes devoted to it in major AI conferences. These initiatives are closer to the concrete down-to-earth approach of Walsh and Mitchell. Another recent book, *Human Compatible* by Stuart Russell⁷ advocates a new research direction in human-machine interaction with control issues at the forefront. As a Google team wrote¹: “one need not invoke ... extreme scenarios to productively discuss accidents, and in fact doing so can lead to unnecessarily speculative discussions that lack precision ... We believe it is usually most productive to frame accident risk in terms of practical (though often quite general) issues with modern ML techniques.”

This harks back to the wise words of Francois Jacob:⁵ “The beginning of modern science can be dated from the time when such general questions as ‘How was the Universe created’ ... ‘What is the essence of Life’ were replaced by more modest questions like ‘How does a stone fall?’ ‘How does water flow in a tube?’ ... While asking very general questions leads to very limited answers, asking limited questions turned out to provide more and more general answers.”

References

1. Amodei, D. et al. Concrete problems in AI safety. arxiv. 2016.
2. Brooks, R. The seven deadly sins of AI predictions. *MIT Technology Review* (Oct. 6, 2017).
3. Domingos, P. *The Master Algorithm*, Basic Books, 2015.
4. Grace, K. et al. When Will AI Exceed Human Performance: Evidence from the Experts. arxiv 2017.
5. Jacob, F. Evolution and tinkering. *Science* 196, 4295 (1977), 1161–1166.
6. Mitchell, M. *Artificial Intelligence: A Guide for Thinking Humans*, Farrar, Straus and Giroux, 2019.
7. Russell, S. *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking, 2019.
8. Tegmark, M. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf, 2017.
9. Walsh, T. *Android Dreams: The Past, the Present and the Future of Artificial Intelligence*. C. Hurst & Co Publishers Ltd, 2017.

Devdatt Dubhashi (dubhashi@chalmers.se) is a professor in the Division of Data Science and AI Department of Computer Science and Engineering at Chalmers University, Sweden.

Copyright held by authors.

^b <https://stanford.io/2VjkiB7>

Article development led by **acmqueue**
queue.acm.org

Creating a software solution with fast decision capability, agile project management, and extreme low-code technology.

BY JOÃO VARAJÃO

Software Development in Disruptive Times

THE RECENT PANDEMIC has brought challenges rarely seen before. It has made evident a world that is strongly globalized, capable, and characterized by a high interdependence of resources and means, but that is also fragile and has a high potential for contamination—not only in the physical sense but also concerning information, ideas, processes, and other aspects.

Given the novelty of the situation, one may be tempted to think this is a unique situation that will soon be overcome, returning eventually to the (apparent) stability that existed previously. However, the reality indicates this view is, at best, illusory and that we live in an age in which societal fragilities and instabilities will be increasingly

evident (optimistically, awareness of them will also become more acute). In other words, crises have always been part of human evolution, and they must be seen as inevitable and recurring realities that need quick and effective responses. The key is to be prepared for them and act accordingly.

As history has so often demonstrated, difficult times enhance society's ability to adapt, and lead to the search for better solutions. Information technologies, which in recent decades have revolutionized the lives of people and businesses—sometimes more or less quietly, sometimes with a bang—are inevitable since they provide cost-effective solutions to the increasingly complex problems of an interconnected and interdependent world. This is easy to understand from a simple example: if, in this pandem-





ic, there had been a global shutdown of the technological infrastructure that supports the Internet, the world would indeed be experiencing a much more complicated and chaotic reality than we are living—and it is already quite difficult for everyone.

It is in this context that the software-development project described in this article is worth reporting on since it involves several disruptive aspects that are fundamental in a world that requires solutions “thought today” to be “made available yesterday.” From the point of market-opportunity awareness to the availability of a fully functional software product, this project took three weeks to complete and involved several state-of-the-art practices and tools: fast decision making; agile project management; and extreme low-code software-development technology.

From Opportunity Awareness to the Decision to Go Forward

On March 12, 2020, the coronavirus outbreak was officially declared a pandemic by the World Health Organization. On the same day, the CEO of Quidgest, a medium-sized software house, identified the pandemic as an opportunity for the company and wrote a plan for creating a web-based software product, together with a preliminary list of requirements (related to monitoring, control, innovation management, and eradication of diseases). This plan was then sent to the sales and marketing teams (both national and international), as well as the research and development teams for feedback. Note that the sales and marketing teams had developed several contacts with government administrations, hospitals, retirement homes, among

others, that provided broad insight into the project’s market potential.

The response from all teams was quick and positive, so the company decided to go forward with the project one day later, on March 13.

Product, Requirements, and Development Milestones

The product, named VIRVI—Health Vigilance and Control Software—is presented as “hyper-agile emergency software for global epidemiologic challenges.” As described by Quidgest, VIRVI is “an information system aimed at supporting the monitoring and control of a virus epidemic, like COVID-19, in any country or region, in an emergency timeframe (that is, starting operating in hours). VIRVI is robust, reliable, and capable of continually evolving, forming the basis for

good critical management and communication facing virus epidemics.”

As depicted in the timeline in accompanying figure, VIRVI evolved quickly. On March 12, the project started with an estimated size of 207 function points, which measure software size defined as the amount of business functionality that an information system (as a product) provides to users; the complexity level was 941 (according to the company’s internal measure that takes into account the number of data tables, data fields, foreign keys, arrays, formulas, forms, form fields, menus, multilanguage, and other aspects, as well as maintenance and documentation requirements). At the product launch three weeks later, the total implemented function points were 1,365, with a complexity level of 4,387.

This illustrates the high volatility of requirements, regularly updated during the project execution following the evolution of the pandemic and the forecasted market opportunities. After looking at national and international best practices and trends, Quidgest decided to focus its solution on nursing and retirement homes (a priority in fighting the pandemic), national health authorities, hospitals, and health laboratories.

The first fully functional prototype of VIRVI was made available on March 22, and in the following week, it was presented in four countries, with three formal demonstrations and two leads generated. This garnered feedback from partners and led to the identification of additional required features. Furthermore, it confirmed that it would be possible to respond quickly to each client’s particular needs and accommodate new requirements.

Some of the final features of the

software are the following: registration and data collection in a centralized information repository; coordination among entities involved in responding to the epidemic, with different access levels; monitoring and decision-making in real time; availability of open data for external analysis; data visualization, with the provision of indicators, statistics, and maps in real time; support for official emergency channels; direct communication between citizens/entities and health professionals; and a centralized marketplace for entities and suppliers.

VIRVI was ready to launch April 3, three weeks after the start of the project. The final product has 117 data tables, 506 data fields, 156 foreign keys, 31 arrays, 115 formulas, 106 forms, 867 form fields, and 274 menus.

Teams, Project Management, and Development Technology

How did Quidgest manage to get a fully functional product in only three weeks, developed in a highly unstable environment with requirements evolving daily and with the described complexity? This section provides some insight.

In the kick-off of the project and during the first two weeks, the team consisted of two full-time system analysts/developers; one part-time (25%) UX (user experience) designer; one full-time project manager; one part-time (50%) development-technology expert; five part-time (25%) people from the marketing team; and four part-time (25%) people from the sales team. After the first two weeks, the team changed to two part-time (50%) system analysts/developers; one part-time (50%) project manager; one part-time (25%) development-technology expert; one part-time (50%) public-health consultant; four

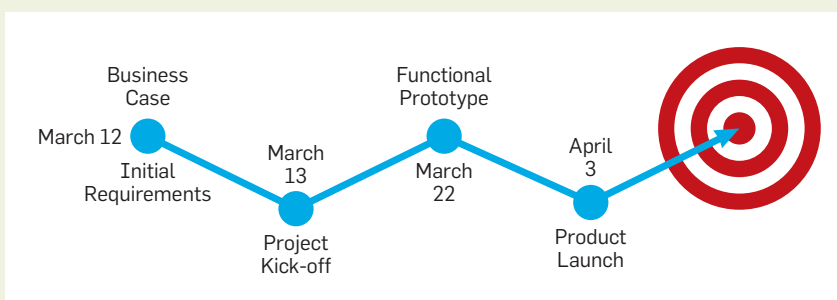
part-time (5%) people from the marketing team; and four part-time (20%) people from the sales team. Taken together, this corresponds to about 3.25 FTEs (full-time equivalents) for the development team and nearly 1.84 FTEs for the marketing and sales teams during the three-week development timeline. Considering the total of 1,365 function points for the final product, the average weekly productivity of the analysts/developers/technology expert was nearly 220 function points.

Since there was no conventional enactment of requirements, the project began based only on the “perception” of what features would be important to the product (based mainly on the CEO’s and the company’s experience in the area, as well as a review of the literature and news). The team also had members with training and expertise in the health sector, and several contacts were made with specialists in health care. Later Quidgest’s health coordinator joined the team, promoting the establishment of a voluntary consultancy scheme. The result of all these voices was a daily evolution of the software requirements. In the end, the actual requirements came from a combination of the company’s experience, contact with specialists, news from the media, scientific studies and official reports on the pandemic.

Given the project’s characteristics, the company’s standard project-management approach was unfeasible in this case since the market window of opportunity was very tight, and the needs of different stakeholders had to be accommodated in the final analysis. Thus, it was necessary to take the agile approach to project management to the extreme because of the need for accepting new requirements almost daily.

The Planner in Microsoft Teams was used to support the process, focusing on teams, tasks, and priorities (established according to the milestones). A Kanban board was created, which could be edited by the whole team, with six states: (1) tasks under analysis, (2) pending, (3) in progress, (4) executed, (5) for testing, and (6) finished. Regular team meetings were held to review the tasks regarding priorities, obsolescence, and clarification of requirements. Another striking aspect of the project was that it was carried out entirely remotely (as of March 16, the company started to work

Project milestones.



remotely due to the pandemic lockdown restrictions).

Another ingredient crucial to the success of the project was the low-code technology adopted for software development. Taking into account the enormous volatility of the requirements and the need to provide fully functional solutions quickly, the challenge was how to achieve this without compromising the quality of the software, the documentation, or further required maintenance.

In this context, the power and usefulness of low-code or extreme low-code technologies become even more evident. These technologies currently include the IBM Automation platform, Zoho Creator, Appian, Mendix, OutSystems, AgilePoint, Google App Maker, Nintex, TrackVia, Quickbase, ServiceNow, Salesforce App Cloud, Microsoft Power Apps, Oracle Visual Builder, and Oracle APEX (Application Express), just to name a few. The distinctive feature of these technologies is their ability to create software applications with minimal hand-coding. For the project, Quidgest used Genio, its proprietary platform. Genio is an extreme low-code (between low-code and no-code) development technology. The development is pattern-oriented, and Genio has code-generation features based on modeling (model-driven engineering).

The chosen technology allowed evolutionary versions of the software to be released on a daily basis. In this way, the development process could be streamlined. Even with the requirements constantly evolving, the next day there could already be a functional solution to support them. The advantages of low-code technology are also fundamental to the maintenance of solutions. The benefits are, perhaps, even more noticeable in this context, by making it possible to support the continuous evolution of requirements without compromising the quality of the software artifacts (including documentation).

Overall, the objectives defined for this project were met despite the high risk caused by the instability of requirements and application scenarios, as well as the urgency of the solution. In a short period of time, Quidgest's project not only gave rise to several sales leads but also contributed to organizational learning and the visibility of the company.



From the point of market-opportunity awareness to the availability of a fully functional software product, the project took three weeks to complete and involved several state-of-the-art practices and tools.



Conclusion

In this project, the challenge was to “deploy software faster than the coronavirus spread.” In a project with such peculiar characteristics, several factors can influence success, but some clearly stand out: top management support, agility (in decision and management), understanding and commitment of the project team, and the technology used. Conventional development approaches and technologies would simply not be able to meet the requirements promptly.

The project described here reflects the demands currently placed on companies in terms of decision and action capacity. It combines market vision and rapid decision-making capacity with action. The company identified an opportunity, defined a project, and decided to move forward, structuring and organizing a team by adopting a different approach to project management—a streamlined agile approach—and adopting a proactive marketing posture. Without technology that supported the rapid development and deployment of software, however, the project could not have been achieved in such a short time—in a context of high instability and rapid evolution of requirements.

A study published by IBM in 2009, “The Enterprise of the Future – Implications for the CIO,” stated, “The enterprise of the future is hungry for change, innovative beyond customer imagination, globally integrated, disruptive by nature, genuine, not just generous.” These are, more than ever, fundamental characteristics for today’s organizations, to which could be explicitly added, “supported by stable as well as disruptive information technology.” Low-code, extreme low-code, and no-code software development, supported by innovative technologies such as artificial intelligence, certainly have influence in this scenario and are expected to accelerate rapidly toward worldwide adoption as major enablers of digital transformation. 

João Varajão is a professor of information systems and project management at the Department of Information Systems of the University of Minho, Portugal. He is also a researcher at the ALGORITMI Research Center. Previously, he worked as an IT/IS consultant, project manager, information systems analyst, and software developer for private companies and public institutions.

Copyright held by author/owner.
Publication rights licensed to ACM.

Article development led by [acmqueue](https://queue.acm.org)
queue.acm.org

Time to move forward from decades-old design.

BY JESSIE FRAZELLE

A New Era for Mechanical CAD

more online

A version of this article with embedded informational links is available at <https://queue.acm.org/detail.cfm?id=3469844>

COMPUTER-AIDED DESIGN (CAD) has been around since the 1950s. The first graphical CAD program, called Sketchpad, came out of MIT (designworldonline.com). Since then, CAD has become essential to designing and manufacturing hardware products. Today,

there are multiple types of CAD. This article focuses on mechanical CAD, used for mechanical engineering.

Digging into the history of computer graphics reveals some interesting connections between the most ambitious and notorious engineers. Ivan Sutherland, who received the Turing Award for Sketchpad in 1988, had Edwin Catmull as a student. Catmull and Pat Hanrahan received the ACM A.M. 2019 Turing Award for their contributions to computer graphics. This included their work at Pixar building RenderMan

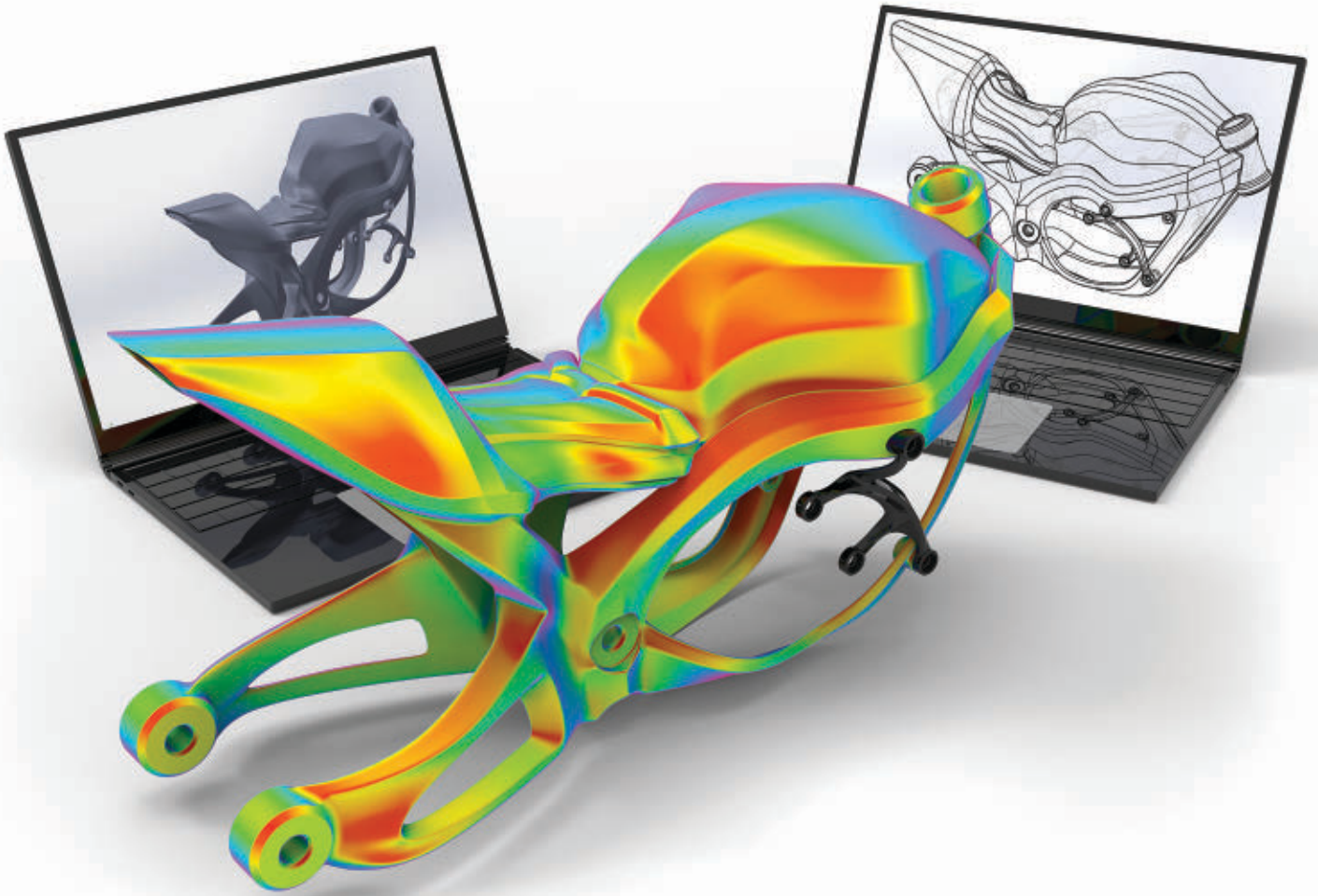
(pixar.com), which was licensed to other filmmakers. This led to innovations in hardware, software, and GPUs. Without these innovators, there would be no mechanical CAD, nor would animated films be as sophisticated as they are today. There would not even be GPUs.

Modeling geometries has evolved greatly over time. Solids were first modeled as wireframes by representing the object by its edges, line curves, and vertices. This evolved into surface representation using faces, surfaces, edges, and vertices. Surface representation is valuable in robot path planning as well. Wireframe and surface representation contains only geometrical data. Today, modeling includes topological information to describe how the object is bounded and connected, and to describe its neighborhood. (A *neighborhood* of a point consists of a set of points containing that point where one can move some distance in any direction away from that point without leaving the set.)

OpenCascade, Parasolid, and ACIS are all boundary-representation (B-rep) kernels. A B-rep model is composed of geometry and topology information. The topology information differs depending on the program used. B-rep file formats include STEP (Standard for the Exchange of Product Model Data), IGES (Initial Graphics Exchange Specification), NX's prt, Solid Edge's par and asm, Creo's prt and asm, SolidWorks' sldprt and sldasm, Inventor's ipt and iam, and AutoCAD's dwg.

Visual representation (vis-rep) models tend to be much smaller in data size than B-rep models. This is because they do not contain as much structural or product management information. Vis-rep models are approximations of geometry and are composed of a mass of flat polygons. Vis-rep file formats include obj, STL, 3D XML, 3D PDF, COLLADA, and PLY.

CAD programs tend to use B-rep models, while animation, game development, augmented reality, and virtual reality tend to use vis-rep models.



However, the two are interchanged frequently. For example, if you were using a B-rep model for manufacturing but wanted to load it into Apple's ARKit for some animations, you would first convert it to COLLADA, a vis-rep file format. The file should already be a bit smaller from dropping all the CAD data, but if you wanted to make it even smaller, you could tweak the polygon counts on each of the meshes for the various parts.

The tools used to build with today are supported on the shoulders of giants, but a lot could be done to make them even better. At some point, mechanical CAD lost some of its roots of innovation. Let's dive into a few of the problems with the CAD programs that exist today and see how to make them better.

Single Threaded

Since most CAD kernels are built on cores from the 1980s, they are not meant for modern systems. Even the latest CPU or GPU won't do much to

help the performance since most of these programs are single threaded, or have single-threaded aspects, and have no awareness of a GPU. OpenSCAD and everything built on CGAL (Computational Geometry Algorithms Library) are single threaded. Sure, some of these kernels have been updated since the 1980s, but their roots are still tied to their predecessors. (I am sure there is a lot to learn from these codebases, but as someone who has seen many old codebases, I know this can lead down a dangerous path.)

This does not mean that all CAD kernels are *entirely* single threaded. Parasolid is multithreaded, but that still means if you are importing or exporting to a file format other than Parasolid, you might have just switched back to a single-threaded process. Another example of a multithreaded kernel is ImplicitCAD (implicitcad.org), which is written in Haskell.

One problem with making a whole CAD program multithreaded is the

different file formats. For example, a STEP file, whose format dates back to the 1980s (iso.org), pretty much mandates the need for a single-threaded process. (Additionally, a STEP file cannot be read sequentially; it must be loaded into memory and then resolved.) Most parametric CAD operations are single threaded; however, the open source project SolveSpace (solvespace.com), which uses NURBS (nonuniform rational basis splines), has some parallel operations.

Duplication

In software development, a pointer is used to get the contents of a memory address. This allows users to reference that same content over and over again without the expense of copying the content itself.

Some products built using CAD may never duplicate a part of their model—lucky for them! For people who do have multiple similar parts in their CAD designs, most CAD programs are creating

very expensive copies of these parts.

For example, imagine a model of a server rack. The *default* method of copying a part (using copy and paste) in SolidWorks, as well as many other industrywide CAD programs, is to copy the entire contents of a child model to a new model. So, if you have 32 sleds in the server rack and use the default copy method in SolidWorks, you have 32 of the exact same model in individual copies. This is very expensive. Each sled has many more models inside, and then those models have child models as well. This exponentially increases the workload on the kernel and on your program to load your model in the first place, since the program does not know these are all the same thing.

Taking a lesson from software development, what you really want is a form of pointer to the model. In the CAD world, these are called *instances*. Then you can have one copy of the model stored, and all the other instances are actually just references to the original copy. This also saves the user a bunch of time. Imagine having to update parts of models in 32 different locations when a part in a sled changes. A wise person once said, “The definition of insanity is doing the same thing over and over again but expecting different results.”

SolidWorks does offer another option that is more in line with how pointers operate, but since this is not the default, most users might not even realize there is a better way. The default path should lead to the least amount of pain. Instead of having two methods for copying, products should have just one. They should make the default method act more like a pointer (or instance) *until* the geometry, surfaces, or topology of the copy (not the main) has changed. In this case the user should be warned that this will now act like a unique part aside from the main copy. Or the user might have mistakenly meant to apply those changes to all the copies, in which case the changes should be applied to the main copy.

There is another huge problem with this. Each CAD program has its own way of implementing and referencing instances. If you export your design from one CAD program to another, you

will likely still have 32 individual sleds rather than one sled and 31 references to the original with only the xyz coordinates changed. Some programs offer ways to import instances, but they all rely on the file format being imported and whether they have the support for that format.

Even if you are using instances, you are still at the mercy of the single-threaded kernel, and none of the copies are likely to render in parallel.

Version Control

For software teams that are accustomed to using git, being able to diff, fix merge conflicts, and work as a team in parallel on the same file is a huge time saver. A number of startups are working to bring this ability to mechanical CAD.

Instead of reinventing version control for CAD, those who use git today want to continue to use git and not have to add another tool to their workflow. Today there is no way to push a CAD file to a git repo, have several people modify the file, and resolve merge conflicts. (Well, maybe it could be done, but it would be the opposite of fun.) For all the startups working to solve version control for mechanical CAD, this is why they had to reinvent the wheel.

In a world where a kernel can fully utilize a modern CPU and GPU, can you not also use a file format that is human-readable and would allow for resolving merge conflicts? When you ask, “What is human-readable and works well with git?” the first answer that comes to mind is a programming language.

The other great win from using a programming language is this: Even if you don’t use or want to use git, there are already many different options for version control of human-readable files. Additionally, integrations with GitHub and other version-control tools could be extended with wasm (WebAssembly) support so that diffs could be visualized as renders as well.

Programmable

Think back to the example of the rack of servers. If part of the rack contained complex math that you were calculating in a program such as Mathematica, you would have to reevaluate the math continuously in another program and update it in your model. If, instead,

you could program in the CAD product itself, then you could do all your calculations in one place and the model would update if anything in the equations changed.

Each sled in the rack of servers has network cables that connect to the back of the sled. Using the GUI, you would have difficulty making these align perfectly with the connector on the sled. Someone would have to sit with the model for an hour or so just tweaking each cable to be perfectly aligned—a huge waste of time. Instead, if you could program the alignment of the cables, you could ensure each was perfectly aligned with the connector.

The need for programming becomes even more acute if you want to do mesh or topology optimizations. Unfortunately, most optimizations are implemented through GUI click interfaces, and given their complexity to define, can often be more trouble than they are worth. Today, some programs allow for scripting, but their APIs are COM (Component Object Model) based and, as you can imagine, built in the 1990s. It’s great they even offer this, however. (Thank you, AutoCAD, for being the first CAD command-line interface I ever used.)

For the modern world, it would be great to generate SDK clients for the CAD program in every language, much the way API clients are generated. This would allow anyone to program in any language. It would lower the barrier to entry since learning a new language would not be required. This would allow for complex math to be done in the CAD program itself rather than using Mathematica, MATLAB, or Wolfram Alpha.

A few scriptable CAD programs exist today and are paving the way for this transition: ImplicitCAD [implicitcad.org], libfive Studio (libfive.com), OpenSCAD (openscad.org), CadQuery (github.com), FreeCAD (freecadweb.org), and ruckus (github.com). Blender (blender.org) has a great console interface. Three.js (threejs.org), while not CAD oriented, is also another great example of a 3D programming language. Jonathan Blow’s Jai (oxide.computer) is for writing systems-level code and a great example of creating a language thinking heavily about performance. (This is not yet open to the public, but

he has talked about it extensively.)

Most of the mechanical engineering community is tied to the GUI, so generating code from GUI interactions would be necessary. This is quite similar to an HTML point-and-click GUI that generates code on the back end. This allows people who want to script to script, and others who want to click can click. Both worlds can be happy—code on the left side, render on the right, just like a markdown editor.

If there is an SDK client for the CAD program and underlying kernel, you can imagine a rich ecosystem of plug-ins and tools emerging, much like the ecosystem that surrounds VSCode, Vim, and Emacs. Most CAD editors used for products are closed off and don't allow for this type of community-based development and sharing. Plug-ins could be written for any use case: for example, mesh/topology optimizations and supply-chain system integrations. This includes the functionality for finding parts, creating BOMs (bills of materials), and computing lead times for parts of the model. Today, this is usually done in separate programs or even spreadsheets.

Plug-ins that support a command+P function would be welcome. In most programs, when you want to print something, you hit command+P. (Creo probably has the closest thing to this functionality but lacks an open ecosystem.) For mechanical CAD, when you want to print your model the underlying program should discover all the 3D printers and machines on the local network (or plugged directly into your machine) and send the parts of your model that are compatible with each machine to be printed. This could even be taken a step further—in a fully automated factory with robots, the program should set up and start the assembly for the model and all the parts.

Speaking of 3D printing, let's look at the STL file format. This format was defined in 1987, and its namesake comes from stereolithography, the first method of additive manufacturing. STL files represent geometry in a series of triangular surfaces. Since STL is a vis-rep format, it does not hold any data about internal structure, color, texture, or any other CAD data that a B-rep format would contain. Modern 3D printers have innovated past the simplicity of the STL format. For example, to print

a full-color model, users need a VRML (Virtual Reality Modeling Language) file, or an STL file associated with textures in order for the printer to add color and texture to the object. Plug-ins can ensure that the printer gets the correct data for the specific model to be printed, without the pain of conversion and ensuring that no materials or textures are dropped.

Testing

The test flow of CAD models usually consists of running simulations. Let's use airflow and thermals as examples.

In the software world, after pushing your code updates, typically a continuous integration (CI) is run on the changes, letting you and your teammates know if you broke anything or if your code is safe to merge. CAD should work the same way. If you make changes to a model, your simulations should run in a CI to let your teammates know if your code is safe to merge. Most of these simulations are compute intensive, so being able to offload the simulations to the cloud or remote servers would also be ideal.

Much as VSCode and other editors have nice plug-ins for offloading tests to other computers, a modern CAD program should have the same.

User Experience and Design

After trying many different industry CAD programs, I have found that most have one characteristic in common: a user interface that looks like it is from the 1990s. It is a bit ironic that a tool used for mechanical design has not considered the design and experience of its user interface. Most CAD programs are in need of a makeover, though there are a couple of outliers that do interface design well: Shapr3D (shapr3d.com), an iPad app, has a great design and intuitive interface; SketchUp (sketchup.com) has an intuitive and beautiful design.

Additionally, CAD applications need to be native on MacOS, Linux, and Windows. Native applications built for their specific platform perform better than ones built with Electron and the like. (That being said, VSCode is a nice Electron app.) Especially for a program as graphics heavy as CAD, using the underlying operating-system graphics mechanisms helps achieve the best

performance possible. Today, a CAD program can be used only on the operating system that is supported by that specific program. Additionally, most use archaic GUI frameworks that truly show their age.

Onshape (onshape.com) changed the mold by offering a SaaS (software-as-a-service) CAD program. This allows expensive compute processes to be easily offloaded to the cloud. This was a truly revolutionary idea, but it limits the user's ability to work offline. In contrast, native apps can work offline but also have the ability to offload workloads to the cloud when connected to the network.

If CAD programs can focus on an intuitive design without falling into a trap of complexity, both new users and professionals should be productive. Just as I would use Vim for side projects as well as professional jobs, I would expect my CAD tool to work just as well for building a toy for fun as it would for a complex project. A lot of this capability comes down to the interface design and extensibility through plug-ins.

A Better Tomorrow

Developers of new CAD programs must think through each of these aspects. No existing CAD program has solved all of these problems.

The world owes so much of the amazing innovation of computer graphics to brilliant people such as Ivan Sutherland, Pat Hanrahan, Ed Catmull, John Carmack, and many others. I can only hope some truly revolutionary changes are headed to the world of computer-aided design in the same way that computer graphics pioneers paved the way for rendering, animations, and virtual reality.

The hardware industry is desperate for a modern way to do mechanical design. A new CAD program created for the modern world would lower the barrier to building hardware, decrease the time of development, and usher in a new era of building. □

Jessie Frazelle is the cofounder and chief product officer of the Oxide Computer Company. Before that, she worked on various parts of Linux, including containers, as well as the Go programming language.

Copyright held by author/owner.
Publication rights licensed to ACM.

[more online](#)

A version of this article with embedded informational links is available at <https://queue.acm.org/detail.cfm?id=3469844>

DOI:10.1145/3445972

Technologies for manipulating and faking online media may outpace people's ability to tell the difference.

BY MATTHEW GROH, ZIV EPSTEIN, NICK OBRADOVICH, MANUEL CEBRIAN, AND IYAD RAHWAN

Human Detection of Machine-Manipulated Media

THE RECENT EMERGENCE of artificial intelligence (AI)-powered media manipulations has widespread societal implications for journalism and democracy,⁷ national security,¹ and art.^{8,14} AI models have the potential to scale misinformation to unprecedented levels by creating various forms of synthetic media.²¹ For example, AI systems can synthesize realistic video portraits of an individual with full control of facial expressions, including eye and lip movement;^{11,18,34–36} clone a speaker's voice with a few training samples and generate new natural-sounding audio of something the speaker never said;² synthesize visually indicated sound effects;²⁸ generate high-quality,

relevant text based on an initial prompt;³¹ produce photorealistic images of a variety of objects from text inputs;^{5,17,27} and generate photorealistic videos of people expressing emotions from only a single image.^{3,40} The technologies for producing machine-generated, fake media online may outpace the ability to manually detect and respond to such media.

We developed a neural network architecture that combines instance segmentation with image inpainting to automatically remove people and other objects from images.^{13,39} Figure 1 presents four examples of participant-submitted images and their transformations. The AI, which we call a “target object removal architecture,” detects an object, removes it, and replaces its pixels with pixels that approximate what the background should look like without the object. This architecture operationalizes one of the oldest forms of media manipulation, known in Latin as *damnatio memoriae*, which means erasing someone from official accounts.

The earliest known instances of *damnatio memoriae* were discovered in ancient Egyptian artifacts, and similar patterns of removal have appeared since.^{10,37} Historically, visual and audio manipulations required both skilled experts and a significant investment of time and resources. Our architecture can produce photo-

» key insights

- **The speed at which misinformation can be produced is faster than it has ever been. By combining instance segmentation with image inpainting, we present an AI model that can automatically and plausibly disappear objects such as people, cars, and dogs from images.**
- **Exposure to manipulated content can prepare people to detect future manipulations. After seeing examples of manipulated images produced by the target object removal architecture, people learn to more accurately discern between manipulated and original images. Participant performance improves more after being exposed to subtle manipulations than blatant ones.**

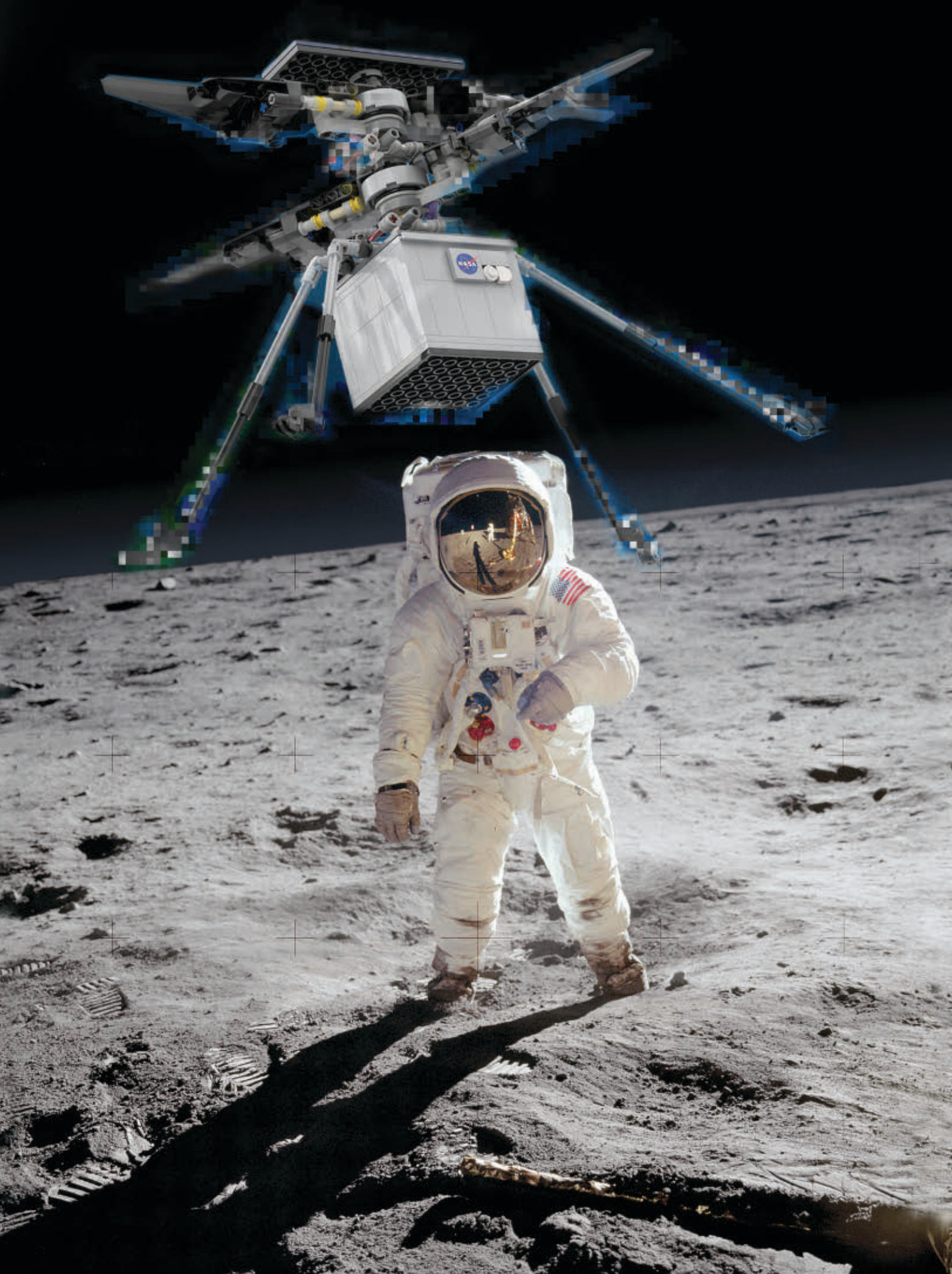


Figure 1. Examples of original images on the top row and manipulated images on the bottom row.



realistic manipulations nearly instantaneously, which magnifies the potential scale of misinformation. This new capacity for scalable manipulation raises the question of how prepared people are to detect manipulated media.

To publicly expose the realism of AI-media manipulations, we hosted a website called Deep Angel, where anyone in the world could examine our neural-network architecture and its resulting manipulations. Between August 2018 and May 2019, 110,000 people visited the website. We integrated a randomized experiment based on a two-alternative, forced-choice design within the Deep Angel website to examine how repeated exposure to machine-manipulated images affects an individual's ability to accurately identify manipulated imagery.

Two-Alternative, Forced-Choice Randomized Experiment

On the Deep Angel website's "Detect Fakes" page, participants are present-

ed with two images consistent with standard two-alternative, forced-choice methodology and are asked a single question: "Which image has something removed by Deep Angel?" The pair of images contains one image manipulated by AI and one unaltered image. After the participant selects an image, the website reveals the manipulation and asks the participant to try again. The MIT Committee on the Use of Humans as Experimental Subjects (COUHES) approved IRB 1807431100 for this study on July 26, 2018.

The manipulated images are drawn from a population of 440 images submitted by participants to be shared publicly. The population of unaltered images contains 5,008 images from the MS-COCO dataset.²³ Images are randomly selected with replacements from each population of images. By randomizing the order of images that participants see, this experiment can causally evaluate the effect of image order on participants' ability to recognize fake media. We test the causal ef-

fects with the following linear probability models:

$$y_{ij} = \mathbf{X}\alpha + \beta \log(T_{i,n,j}) + \mu_i + \nu_j + \epsilon_{i,j,n} \quad (1)$$

and

$$y_{ij} = \mathbf{X}\alpha + \beta_1 T_{i_1,j} + \beta_2 T_{i_2,j} + \dots + \beta_{10} T_{i_{10},j} + \mu_i + \nu_j + \epsilon_{i,j,n} \quad (2)$$

where y_{ij} is the binary accuracy (correct or incorrect guess) of participant j on manipulated image i . X is a matrix of covariates indexed by i and j , T_{i_n} represents the order n in which manipulated image i appears to participant j , μ_i represents the manipulated image-fixed effects, ν_j represents the participant-fixed effects, and $\epsilon_{i,j}$ represents the error term. The first model fits a logarithmic transformation of T_{i_n} to y_{ij} . The second model estimates treatment effects separately for each image position. Both models use Huber-White (robust) standard errors, and errors are clustered at the image level.

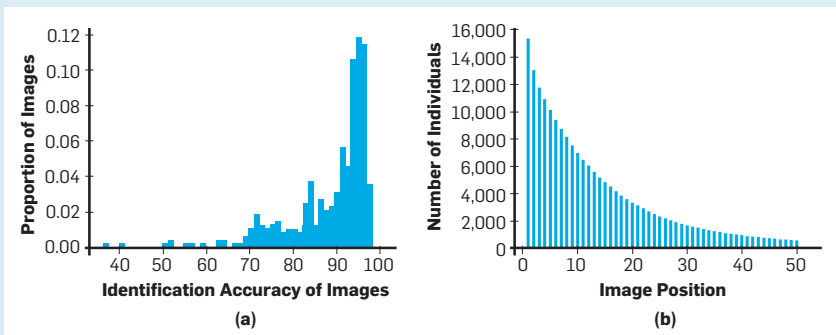
Results

Participation and average accuracy.

From August 2018 to May 2019, 242,216 guesses were submitted from 16,542 unique IP addresses with a mean identification accuracy of 86%. The website did not require participant sign-in, so we study participant behavior under the assumption that each IP address represents a unique individual. The majority of participants participated in the two-alternative, forced-choice experiment multiple times, and 7,576 participants submitted at least 10 guesses.

Each image appears as the first image an average of 35 times and the

Figure 2. (a) Histogram of mean identification accuracies by participants per image (b) Bar chart plotting number of individuals over image position.



tenth image an average of 15 times. The majority of manipulated images were identified correctly more than 90% of the time. In the sample of participants who saw at least 10 images, the mean percentage correct classification is 78% on the first image seen and 88% on the tenth image seen. Figure 2a shows the distribution of identification accuracy across images, while Figure 2b shows the distribution of how many images each participant saw. The interquartile range of the number of guesses per participant is from three to 18 with a median of eight.

Figure 3a plots participant accuracy on the y-axis and image order on the x-axis, revealing a logarithmic relationship between accuracy and exposure to manipulated images. In this plot showing scores for all participants, accuracy increases rapidly over the first 10 images and plateaus around 88%.

Learning rate. With 242,216 observations, we run an ordinary least-squares regression with participant- and image-fixed effects on the likelihood of correctly guessing the manipulated image. The results of these regressions are presented in Tables 1 and 2 in the online appendix (<https://dl.acm.org/doi/10.1145/3445972>). Each column in Tables 1 and 2 adds an incremental filter to offer a series of robustness checks. The first column shows all observations. The second column drops all participants who submitted fewer than 10 guesses and removes all control images where nothing was removed. The third column drops all observations where a participant has already seen an image. The fourth column drops all images qualitatively judged as below very high quality.

Across all four robustness checks with and without fixed-effects, our models show a positive and statistically significant relationship between T_n and $\hat{y}_{i,j}$. In the linear-log model, a one-unit increase in $\log(T_n)$ is associated with a 3% increase in $\hat{y}_{i,j}$. This effect is significant at the $p < .01$ level. In the model that estimates Equation 2, we find a 1% average marginal treatment effect size of image position on $\hat{y}_{i,j}$. This effect is also significant at the $p < .01$ level. In other words, participants improve their ability to guess by 1% for each of the first 10 guesses. Figure 3b shows these results graphically.

Within the context of object removal manipulations, exposure to media manipulation and feedback on what has been manipulated improves a participant's ability to recognize faked media. After getting feedback on 10 pairs of images for an average of 1 min., 14 sec., a participant's ability to detect manipulations improves by 10%. With clear evidence that human detection of machine-manipulated media can improve, the next question is: what is the mechanism that drives participant learning rates? How do feedback, image characteristics, and participant qualities affect learning rates?

Potential Explanatory Mechanisms

We can explore what drives the learning rate by examining heterogeneous effects of image characteristics and participant qualities. Figure 4 presents 10 plots of heterogeneous learning rates based on image-fixed effects regressions with errors clustered at the participant level.

We evaluate the quality of a manipulation across five measures: (a) a subjective quality rating, (b) 1st and 4th quartile image entropy, (c) 1st and 4th quartile proportion of area of the manipulated image, (d) 1st and 4th quartile mean identification accuracy per image, and (e) number of objects disappeared. The subjective quality rating is based on ratings provided by an outside party and is a binary rating based

on whether obvious artifacts were created by the image manipulation.

Image entropy is measured based on delentropy, an extension of Shannon entropy for images.²⁰ To help understand delentropy, Figure 4 presents three pairs of images subjectively rated as high quality. Their corresponding entropy scores are included, along with the proportion of the image transformed, mean accuracy of participants' first guesses, and mean accuracy of subsequent participant guesses to exemplify what study participants learned.

For images subjectively marked as high quality, participants correctly discern 75% for the first image and 83% for the tenth image seen. In contrast, participant accuracy on the low-quality images is higher, at 82% and 94% for the first and tenth image seen, respectively. Table 3 (see online appendix) shows that the difference in means across the subjective quality measure is statistically significant at the 99% confidence level ($p < .01$), but we do not find a statistically significant difference in learning rates.

As seen in Figure 4a, there is evidence that participants learn to identify low-quality images faster than high-quality images if only looking at the first five images seen. When examining the first 10 images seen, we do not find a statistically significant difference in the interaction between subjective quality and the logarithm of the image position. These results indicate that the

Figure 3. Participants' overall and marginal accuracy by image order.

Error bars show a 95% confidence interval for each image position:
 (a) overall accuracy for all participants with no fixed effects
 (b) marginal accuracy (relative to the first image position) for all participants who saw at least 10 images controlling for participant- and image-fixed effects and clustering errors at the image level.
 In (b), the 11th position includes all image positions beyond the 10th.

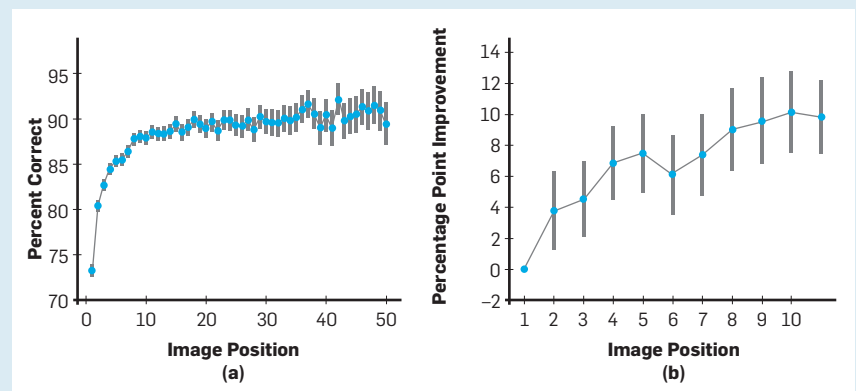
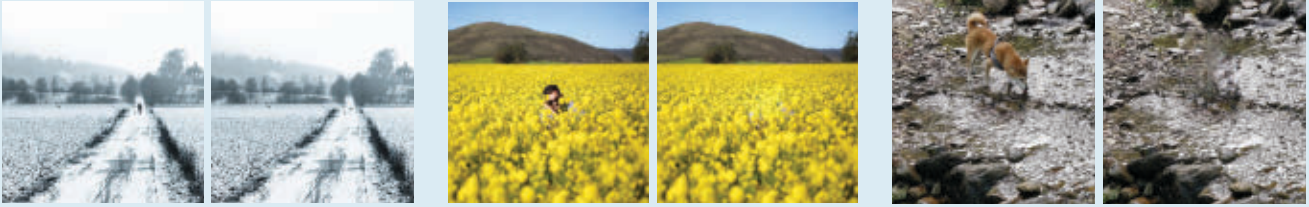
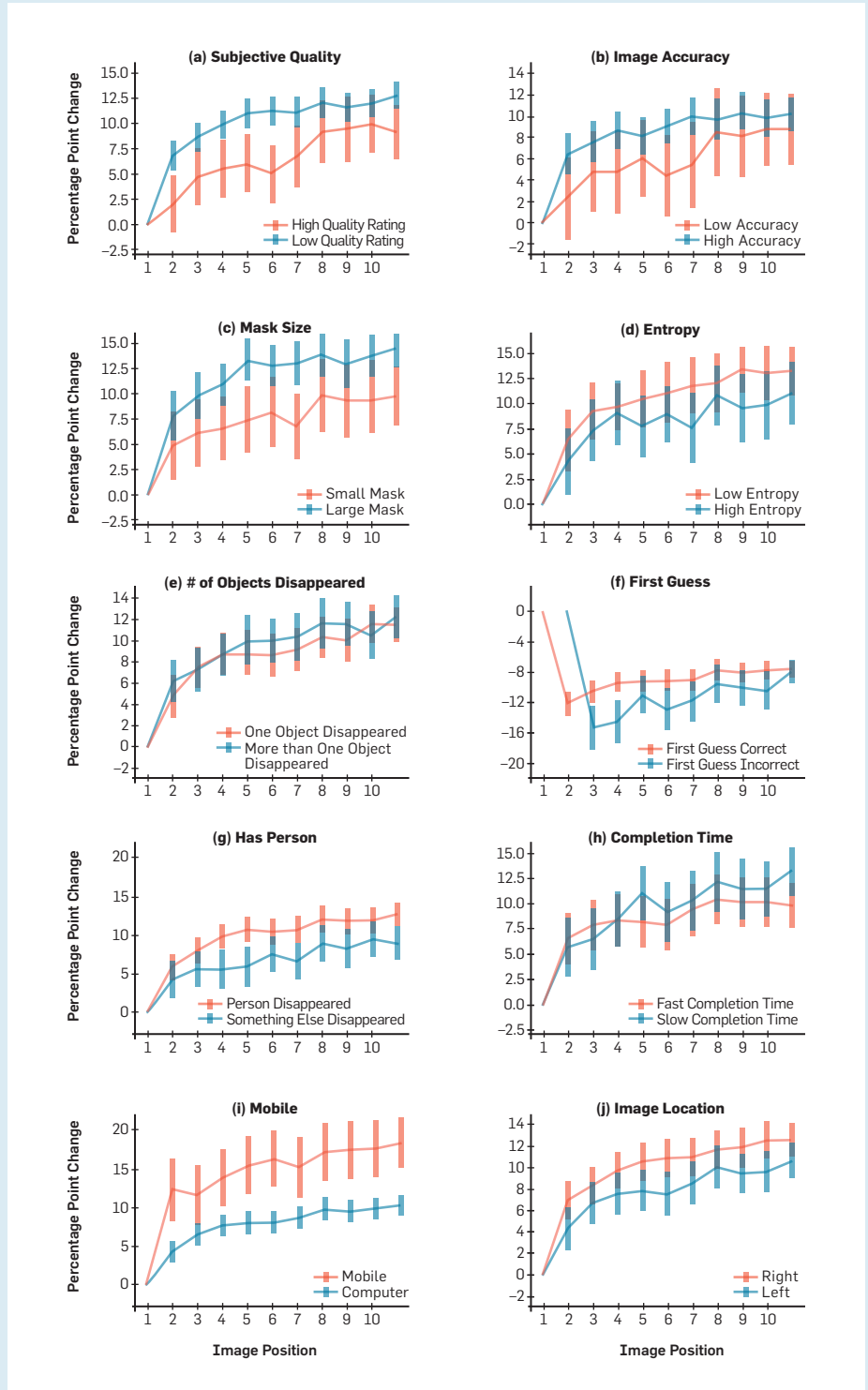


Figure 4. Ten plots of heterogeneous learning rates based on image-fixed effects regressions.



(i) From left to right, images (cropped into squares for display purposes) are increasing in entropy (3.5, 6.0, and 8.5), varying in percent of the image transformed (0.7%, 2.4%, and 1.9%), similar in accuracy on the first guess (70%, 71%, 70%), and varying in accuracy beyond the first guess (88%, 82%, 92%).

(ii) Ten plots display heterogeneous effects of image and participant characteristics on learning while controlling for participant- and image-fixed effects (a) whether the subjective image quality was judged as high by a third party, (b) whether the original image was in the 1st to 25th percentile of accuracy or 75th to 99th, (c) whether the original image was in the 1st to 25th percentile of image mask proportion or 75th to 99th, (d) whether the original image was in the 1st to 25th percentile of entropy or 75th to 99th, (e) whether there were one or multiple objects disappeared (f) whether the participant's first answer was correct (the omitted position for each learning curve represents perfect accuracy), (g) whether the image contained a person, (h) whether the original image was in the 1st to 25th percentile of time to evaluate 10 images or 75th to 99th, (i) whether the participant viewed the images on a mobile device or computer, and (j) whether the image was placed on the left or right side of the screen. The error bars represent the 95% confidence interval for each image position and errors are clustered at the image level.



main effect is not simply driven by participants becoming proficient at guessing low-quality images in our data.

The other proxies for image quality provide insight into how subtleties play a role in discerning image manipulations. Participants learn to identify low-entropy images faster than high-entropy images, and they recognize images with a large masked area faster than images with a small masked area. Table 3 shows that this difference in learning rates is statistically significant at the 95% ($p < .05$) and 90% ($p < .10$) levels, respectively.

Smaller masked areas and lower entropy is associated with less stark and more subtle changes between original and manipulated images. This relationship may indicate that participants learn more from subtle changes than more obvious manipulations. It may even mean people are learning to detect which kinds of images are hard to discern and, therefore, potentially likely to contain a manipulation when no obvious manipulation is apparent. It is important to note that neither the split between the 1st and 4th quartiles of mean accuracy per image nor the split between one object and many disappeared objects has a statistically significant effect on the learning rates. This means we find no association between overall manipulation discernment difficulty and learning rates.


A participant's initial performance is indicative of his or her future performance. In Figure 4, we compare subsequent learning rates of participants who correctly identified a manipulation on their first attempt to participants who failed on their first attempt and succeeded on their second. In this comparison, the omitted position for each learning curve represents perfect accuracy, which makes the marginal effects of subsequent image positions negative relative to these omitted image positions. On the first three of four image positions in this comparison, which correspond to the third through sixth image positions, we find that initially successful participants learn faster than participants who were initially unsuccessful. This heterogeneous effect does not persist in the seventh position or beyond. Overall, this heterogeneous effect is statistically significant at the 99% level ($p < .01$), suggesting that people who are better at discerning manipulations are also faster at learning to discern manipulations.

cerning manipulations are also faster at learning to discern manipulations.


We find participants learn to discern manipulations involving disappeared people faster than images with any other object removed. This difference is statistically significant at the 95% confidence interval ($p < .05$) in the log-linear regression as shown in Table 3. Figure 4 also shows this difference as statistically significant in two of the 10 image positions, suggesting that participants may be learning to detect the kinds of images that are conducive to plausible object removals.

There is a clear difference in the learning rate of participants based on whether they participated with mobile phones or computers. Participants on mobile phones learn at a consistently faster rate than participants on computers, and this difference is statistically significant as shown in Table 3 and displayed across nine of 10 image positions in Figure 4. It is possible that the seamlessness of the zoom feature on a phone relative to a computer enables mobile participants to inspect each image more closely. We do not find evidence that image placement on the website correlates with overall accuracy.

No strong evidence suggests that the speed with which a participant rated 11 images is related to the learning rate, but we do find evidence of an interaction between answering speed upon wrong guesses of high-quality images. In Table 4 (see online appendix), we present a regression of current and lagged features on participant accuracy. It is important to note that we find high-quality images reduce participant accuracy by 4%, which is significant at the 99% confidence interval ($p < .01$), but we do not find a relationship between whether the previous image was high quality and participant accuracy on the current image. However, the interaction of seconds, guessing the previous answer incorrectly, and the previous image being high quality, is associated with a 0.3% increase in participant accuracy for every marginal second ($p < .05$). This correlational evidence suggests that when participants slow down after guessing incorrectly on high-quality, harder-to-guess images, they perform better.



This new capacity for scalable manipulation raises the question of how prepared people are to detect manipulated media.




Discussion


While AI models can improve clinical diagnoses^{9,19,30} and enable autonomous driving,⁶ they also have the potential to scale censorship,³² amplify polarization,⁴ and spread fake news and manipulated media.³⁸ We present results from a large-scale, randomized experiment showing that the combination of exposure to manipulated media and feedback on which media has been manipulated improves an individual's ability to detect media manipulations.

Direct interaction with cutting-edge technologies for content creation might enable more discerning media consumption across society. In practice, the news media has exposed high-profile, AI-manipulated media, including fake videos of the Speaker of the House of Representatives Nancy Pelosi and Facebook CEO Mark Zuckerberg, which serves as feedback to everyone on what manipulations look like.^{24,25} Our results build on recent research showing that people can detect low-quality news,²⁹ human intuition can be a reliable source of information about adversarial perturbations to images,⁴² and familiarizing people with how fake news is produced may confer them with cognitive immunity when they are later exposed to misinformation.³³ Our results also offer suggestive evidence for what drives learning to detect fake content. In this experiment, presenting participants with low-entropy images with minor manipulations on mobile devices increased learning rates at statistically significant levels. Participants appear to learn best from the most subtle manipulations.

Our results focus on a bespoke, custom-designed, neural-network architecture in a controlled, two-alternative, forced-choice experimental setting. The external validity of our findings should be further explored in different domains, using different generative models, and in settings where people are not instructed explicitly to look out for fakes, but rather encounter them in a more naturalistic social media feed, and in the context of reduced attention span. Likewise, future research in human perception of manipulated media should explore to what degree an individual's ability to adaptively detect manipulated media comes from learning by doing, direct feedback, and awareness that any-



With clear evidence that human detection of machine-manipulated media can improve, what is the mechanism that drives participants' learning rates?



thing is manipulated at all.

Our results suggest a need to re-examine the precautionary principle that is commonly applied to content-generation technologies. In 2018, Google published BigGAN, which can generate realistic-appearing objects in images, but while the company hosted the generator for anyone to explore, it explicitly withheld the discriminator for its model.⁵ Similarly, OpenAI restricted access to its GPT-2 model, which can generate plausible long-form stories given an initial text prompt, by only providing a pared-down model of GPT-2 trained with fewer parameters.³¹ If exposure to manipulated content can prepare people to detect future manipulations, then censoring dissemination of AI research on content generation may prove harmful to society by leaving it unprepared for a future of ubiquitous AI-mediated content.

Methods

We developed a *Target Object Removal* architecture, combining instance segmentation with image inpainting to remove objects in images and replace those objects with a plausible background. Technically, we combine a convolutional neural network (CNN) trained to detect objects with a generative adversarial network (GAN) trained to inpaint missing pixels in an image.^{12,13,16,22} Specifically, we generate object masks with a CNN based on a RoIAlign bilinear interpolation on nearby points in the feature map.¹³ We crop the object masks from the image and apply a generative inpainting architecture to fill in the object masks.^{15,39} The generative inpainting architecture is based on dilated CNNs with an adversarial loss function, allowing the generative inpainting architecture to learn semantic information from large-scale datasets and generate missing content that makes contextual sense in the masked portion of the image.³⁹

Target Object Removal Pipeline

Our end-to-end, targeted object removal pipeline consists of three interfacing neural networks:

- **Object Mask Generator (G):** This network creates a segmentation mask $X' = G(X, y)$ given an input image X and a target class y . In our experiments, we initialize **G** from a semantic

segmentation network trained on the 2014 MS-COCO dataset following the Mask-RCNN algorithm.¹³ The network generates masks for all object classes present in an image, and we select only the correct masks based on input y . This network was trained on 60 object classes.

► **Generative Inpainter (I):** This network creates an inpainted version $Z = I(X', X)$ of the input image X and the object mask X' . **I** is initialized following the DeepFill algorithm trained on the MIT Places 2 dataset.^{39,41}


► **Local Discriminator (D):** The final discriminator network takes in the inpainted image and determines its validity. Following the training of a GAN discriminator, **D** is trained simultaneously on **I**, where X are images from the MIT Places 2 dataset and X' are the same images with randomly assigned holes following.^{39,41}

Live Deployment

The Deep Angel website enabled us to make the Target Object Removal architecture publicly available.^a We hosted the architecture API with a single Nvidia Geforce GTX Titan X; anyone could upload an image to the site and select an object to be removed from the image.

Participants uploaded 18,152 unique images from mobile phones and computers; they also directed the crawling of 12,580 unique images from Instagram. We can surface the most plausible object removal manipulations by examining the images with the lowest guessing accuracy. The Target Object Removal architecture can produce plausible content, but the plausibility is largely image dependent and constrained to specific domains, where objects are a small portion of the image, and the background is natural and uncluttered by other objects.

Data availability: The data and replication code are available at: <https://github.com/mattgroh/human-detection-machine-manipulated-media-data-code>.

Acknowledgments. We thank Abhimanyu Dubey, Mohit Tiwari, and David McKenzie for their helpful comments and feedback. 

a We retained the Cyberlaw Clinic from Harvard Law School and Berkman Klein Center for Internet & Society for advice on copyright protection of manipulated images.

References

1. Allen, G. and Chan, T. Artificial intelligence and national security. *Belfer Center for Science and International Affairs*, Cambridge, MA (2017).
2. Arik, S.O., Chen, J., Peng, K., Ping, W., and Zhou, Y. Neural voice cloning with a few samples. *arXiv preprint arXiv:1802.06006* (2018).
3. Averbuch-Elor, H., Cohen-Or, D., Kopf, J., and Cohen, M.F. Bringing portraits to life. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 196.
4. Bakshy, E., Messing, S., and Adamic, L.A. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.
5. Brock, A., Donahue, J., and Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096* (2018).
6. Chen, C., Seff, A., Kornhauser, A., and Xiao, J. Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proc. of the IEEE Intern. Conf. on Computer Vision* (2015), 2722–2730.
7. Chesney, R. and Citron, D.K. Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review* 107, 6 (Dec. 2019).
8. Epstein, Z., Boulais, O., Gordon, S., and Groh, M. Interpolating GANs to scaffold autotelic creativity. *arXiv preprint arXiv:2007.11119* (2020).
9. Esteve, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (2017), 115.
10. Freedberg, D. The power of images: Studies in the history and theory of response. University of Chicago Press (1989).
11. Garrido, P., Valgaerts, L., Sarmadi, H., Steiner, J., Varanasi, K., Perez, P., and Theobalt, C. Vdub: Modifying face video of actors for plausible visual alignment to a dubbed audio track. *Computer Graphics Forum* 34. Wiley Online Library (2015), 193–204.
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. (2014), 2672–2680.
13. He, K., Gkioxari, G., Dollár, P., and Girshick, R.B. Mask R-CNN. *CoRR abs/1703.06870* (2017).
14. Hertzmann, A. Can computers create art? In *Arts 7, Multidisciplinary Digital Publishing Institute* (2018), 18.
15. Iizuka, S., Simo-Serra, E., and Ishikawa, H. Globally and locally consistent image completion. *ACM Trans. on Graphics (Proc. of SIGGRAPH 2017)* 36, 4 (2017).
16. Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).
17. Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. *arXiv preprint arXiv:1812.04948* (2018).
18. Kim, H., Garrido, P., Tewari, A., Xu, W., Thies, J., Nießner, M., Pérez, P., Richardt, C., Zollhöfer, M., and Theobalt, C. Deep video portraits. *arXiv preprint arXiv:1805.11714* (2018).
19. Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., den Heeten, A., and Karssemeijer, N. Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis* 35 (2017), 303–312.
20. Larkin, K.G. Reflections on Shannon information: In search of a natural information-entropy for images. *arXiv preprint arXiv:1609.01117* (2016).
21. Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., et al. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.
22. LeCun, Y., Bengio, Y., and Hinton, G.E. Deep learning. *Nature* 521, 7553 (2015), 436–444.
23. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C.L. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, Springer. (2014), 740–755.
24. Mervosh, S. Distorted videos of Nancy Pelosi spread on Facebook and Twitter, helped by Trump. (May 2019). <https://www.nytimes.com/2019/05/24/us/politics/pelosi-doctored-video.html>.
25. Metz, C. A fake Zuckerberg video challenges Facebook's rules. (June 2019). <https://www.nytimes.com/2019/06/11/technology/fake-zuckerberg-video-facebook.html>.
26. Molodetskikh, I., Erofeev, M., and Vatolin, D. Perceptually motivated method for image inpainting comparison. *arXiv preprint arXiv:1907.06296* (2019).
27. Nguyen, A., Yosinski, J., Bengio, Y., Dosovitskiy, A., and Clune, J. Plug & play generative networks: Conditional iterative generation of images in latent space. *CoRR abs/1612.00005* (2016).

28. Owens, A., Isola, P., McDermott, J.H., Torralba, A., Adelson, E.H., and Freeman, W.T. Visually indicated sounds. *CoRR abs/1512.08512* (2015).
29. Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D., and Rand, D.G. Understanding and reducing the spread of misinformation online. (2019).
30. Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., and Webster, D.R. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* 2, 3 (2018), 158.
31. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. Language models are unsupervised multitask learners. (2019).
32. Roberts, M.E. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press (2018).
33. Roozenbeek, J. and van der Linden, S. Fake news game confers psychological resistance against online misinformation. *Palgrave Communications* 5 (2019).
34. Saito, S., Wei, L., Hu, L., Nagano, K., and Li, H. Photorealistic facial texture inference using deep neural networks. *CoRR abs/1612.00523* (2016).
35. Suwajanakorn, S., Seitz, S.M., and Kemelmacher-Shlizerman, I. Synthesizing Obama: Learning lip sync from audio. *ACM Trans. on Graphics (TOG)* 36, 4 (2017), 95.
36. Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., and Nießner, M. Face2face: Real-time face capture and reenactment of RGB videos. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, (2016), 2387–2395.
37. Varner, E.R. Mutilation and transformation: Damnatio memoriae and Roman imperial portraiture. *Monumenta Graeca et Romana* 10, Brill (2004).
38. Vosoughi, S., Roy, D., and Aral, S. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
39. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T.S. Generative image inpainting with contextual attention. *arXiv preprint arXiv:1801.07892* (2018).
40. Zakharov, E., Shysheya, A., Burkov, E., and Lempitsky, V. Few-shot adversarial learning of realistic neural talking head models (2019).
41. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).
42. Zhou, Z., and Firestone, C. Humans can decipher adversarial images. *Nature Communications* 10, 1 (2019), 1334.

Matthew Groh is a Ph.D. candidate in the Massachusetts Institute of Technology (MIT) Media Lab, Cambridge, MA, USA.


Ziv Epstein is a Ph.D. candidate in the Massachusetts Institute of Technology (MIT) Media Lab, Cambridge, MA, USA.

Nick Obradovich is a senior research scientist and principal investigator in the Center for Humans & Machines at Max Planck Institute for Human Development, Berlin, Germany.

Manuel Cebrian is the Max Planck Research Group Leader of the Digital Mobilization Research Group in the Center for Humans & Machines at Max Planck Institute for Human Development, Berlin, Germany.

Iyad Rahwan is director in the Center for Humans & Machines at Max Planck Institute for Human Development, Berlin, Germany.

Author contributions. M.G. implemented the methods, M.G., Z.E., N.O. analyzed data and wrote the article. All authors conceived the original idea, designed the research, and provided critical feedback on the analysis and manuscript.

 This work is licensed under a <http://creativecommons.org/licenses/by/4.0/>



Watch the authors discuss this work in the exclusive *Communications* video. <https://caom.acm.org/videos/machine-manipulated-media>

DOI:10.1145/3440868

Beyond the pandemic, organizations need to recognize what digital assets, interactions, and communication processes reap the most benefits from virtual reality.

BY OSKU TORRO, HENRI JALO, AND HENRI PIRKKALAINEN

Six Reasons Why Virtual Reality Is a Game-Changing Computing and Communication Platform for Organizations

THE COVID-19 PANDEMIC created unprecedented disruptions to businesses, forcing them to take their activities into the virtual sphere. At the same time, the limitations of remote working tools have become painfully obvious, especially in terms of sustaining task-related focus, creativity, innovation, and

social relations. Some researchers are predicting that the lack of face-to-face communication may lead to decreased economic growth and significant productivity pitfalls in many organizations for years to come.¹³

As the length and lasting effects of the COVID-19 pandemic cannot be reliably estimated, organizations will likely face mounting challenges in the ways they handle remote work practices. Therefore, it is important for organizations to examine which solutions provide the most value in these exceptional times. In this article, we propose virtual reality (VR) as a critical, novel technology that can transform how organizations conduct their operations.

VR technology provides “the effect of immersion in an interactive, three-dimensional, computer-generated environment in which virtual objects have spatial presence.”⁵ VR’s unique potential to foster human cognitive functions (that is, the ability to acquire and pro-





cess information, focus attention, and perform tasks) in simulated environments has been known for decades.^{6,8,32} VR has, thus, long held promise for transforming how we work.³³

Earlier organizational experiments with desktop-based virtual worlds (VWs)—3D worlds that are used via 2D displays—have mostly failed to attract participation and engagement.^{34,37} Increasing sensory immersion has been identified as necessary for mitigating these problems in the future.¹⁸ Therefore, sensory immersion in VR through the use of head-mounted displays (HMDs) can be seen as a significant step forward for organizations transferring their activities to virtual environments. In this regard, VR is now starting to fulfill the expectations that were placed upon VWs in the past decades, as per Benford et al, for instance.⁴

However, VR has only recently matured to a stage where it can truly be

said to have significant potential for wider organizational use.¹⁷ In 2015, Facebook founder and CEO Mark Zuckerberg described VR as “the next major computing and communication platform.”³⁸ Although VR has received this kind of significant commercial attention, its potential in organizational use remains largely scattered or unexplored in the extant scientific literature.

Drawing on contemporary research and practice-driven insights, this article provides six reasons why VR is a fundamentally unique and transformative computing and communication platform that extends the ways organizations use, process, and communicate information. We relate the first three reasons with VR as a computing platform and its potential to foster organizations’ knowledge management processes and the last three reasons with VR as a communication platform and its potential to foster organizations’ remote communication processes.

VR as a Computing Platform: Transformative Knowledge Management

VR can be used to simulate many organizational activities, depending on an organization’s goals and demands. However, VR can also be seen as a transformative knowledge manage-

» key insights

- **VR can solve many critical bottlenecks of conventional remote work while also enabling completely new business opportunities.**
- **VR enables novel knowledge-management practices for organizations via enriched data and information, immersive workflows, and integration with appropriate IS and other emerging technologies.**
- **VR enables high-performing remote communication and collaboration by simulating or transforming organizational communication, in which altered group dynamics and AI agents can also play an interesting role.**

ment system because it provides new ways to manage and enrich information and workflows, and it has significant potential as a platform for integrating other information systems (IS) and emerging technologies. Next, we articulate three reasons why VR is a game-changing computing platform.

Reason 1: Enriched data and information

The current methods for examining complex information via 2D displays impose obvious limitations on the presentation of information to users. For example, it is difficult for users to understand how a certain room layout might fit with their work tasks purely from architectural 2D drawings.¹⁷ VR tackles this problem by enabling enhanced spatial understanding of 3D content and data when compared to traditional 2D displays.^{6,27} In VR, users can examine immersive 3D content spatially from multiple perspectives, such as birds-eye view or 1:1 scale).

In general, the ability to view 3D content in an immersive 3D environment is a powerful tool for fostering users' understanding of complex issues and scenarios.⁸ Users can immerse themselves in the virtual content, which can be anything from the molecular structure of a medicine or the design of a movie scene.¹² In comparison with 2D displays, the information in VR is perceived to be more real and explicit and, thus, less abstract and ambiguous. This has far-reaching consequences for many organizations across different fields.³³

VR technology is also highly adaptable, allowing different layers of information about the same content to be shown according to users' needs or preferences.³⁵ For example, in a virtual building, an architect can work with a different layer of information than a construction engineer or a potential customer. Ideally, this requires the addition of relevant metadata to the digital content to present it to various stakeholders automatically and efficiently on the basis of user profiles. If needed, adaptations in VR can further be based on natural and intuitive user behaviors, such as gaze or body movements.³³ As individuals are able to immerse themselves in data and information, and increase their contextual cognitive func-

tions, VR adaptability has the potential for organizations to foster stakeholder engagement and participation.

Information can also be stored, organized, and retrieved spatially in VR. Spatial awareness (for instance, viewing the world in 3D) has long been used to enhance our information-recall skills. For example, multiple 2D displays, such as virtual desktops or whiteboards, can be positioned to a virtual space in an organized manner to display vast amounts of information.¹⁹ Thus, users, especially in knowledge-intensive work, can personalize their own spatial information management system and increase their productivity through better recall of relevant information.

Reason 2: Immersive workflows and training

Many work activities are still bound to a specific physical space, which can be especially inefficient when large amounts of complex information and multiple stakeholders are involved. Moreover, many organizations still rely on labor-intensive business processes that do not scale efficiently, such as building expensive physical prototypes during product design. For example, if a physical miniature model of a building or a vehicle is created, it can be only displayed at a certain location and at previously agreed-upon times. VR provides an ideal platform for scaling up many of these activities by enabling an organization's stakeholders to manipulate different digital assets directly in VR from anywhere in the world in a shared immersive environment. For example, existing physical assets can be replicated in VR as digital twins to support many different use cases and workflows relating to product development or training.¹⁵

VR's most obvious use cases have long been in different training scenarios, for example, for fire safety or surgeries.³³ These use cases undoubtedly have benefits, especially when substituting activities that are extremely dangerous or expensive.¹⁵ VR provides a major advantage for virtual workflows and training because, in addition to the benefits of enriched data and information, users can have intuitive and natural interactions with the digital content. Mounting evidence over the past three

decades shows that when the VR system realistically responds to the user's actions, the user is likely to react and interact realistically as well.^{32,33} Furthermore, as users perceive training in VR as real, the benefits of VR apply not only in the practice of hard but also soft skills, such as customer engagement or public speaking.³ Therefore, acquiring professional skills and knowledge via the use of VR holds exceptional potential when compared to many conventional IT technologies.

However, VR is not limited to experiences that imitate our real-world expectations. It can also simulate impossible interactions, such as teleportation and moving heavy objects without gravity. VR can, thus, be used to create experiences that are "better than reality,"²¹ based on the desired organizational effect. Organizations can further improve performance by enhancing the user-flow experience and motivation to efficiently perform tasks by gamifying features of VR and aspects of work routines. The user's performance and progression in, for example, different training scenarios can be tracked and verified automatically as in many games. In the context of workflows, for instance, relevant changes in a virtual building can be presented to users with navigation and distance markers or with estimations about changes in costs and the construction schedule.

Another advantage of VR is that it becomes a living 3D document and a version-control system that is modified by user interactions. The information can persist in the virtual environment as long as needed. VR content can be made available anywhere in the world at all times, which enables far more iterative collaboration and knowledge transfer within projects.¹⁷ Users can also return to the digital assets even years after they were last used if they, for example, need to learn how some earlier design challenge was solved. The superior spatial recall of information in VR can further increase user efficiency in these work tasks.¹⁹

Reason 3: Increasing synergies with other emerging technologies and organizational IS

Fluent information transfer between an organization's IS and its stakeholders is critical to the organization's success. Taking into account VR's

capability to enrich information and workflows, using VR as a platform for integrating existing IS comes with many interesting synergies.

For example, architecture, engineering, and construction (AEC) professionals use building information modeling (BIM) as a process to manage all information relating to construction projects. BIM consists not only of the physical 3D characteristics of buildings and infrastructure but also vast amounts of other information, such as construction times, costs, energy performance, and safety aspects. Exporting complex 3D assets, such as BIM, to VR was earlier a limitation in many organizational settings, but the latest VR software has tackled many of these challenges, even enabling live editing of 3D models in VR.²³ As VR can host complex 3D information in an immersive and interactive fashion, integrating organizational digital content, such as BIM, with VR can foster the effectiveness of organizational decision-making and virtual workflows.

It is also important to ensure that the information processed in VR is transferred in the other direction as well (that is, back to relevant IS or software). For example, when a client makes a purchase decision in VR, this information should be directly imported to the customer relationship management (CRM) and enterprise resource planning (ERP) systems. This also eliminates the need to manually edit the assets outside of VR, which reduces mistakes and redundant work. Ideally, the feedback that is given in VR should also provide immediately actionable tasks in other systems. For instance, 3D model annotations in VR should translate to tasks in the design software.

VR has countless technological synergies with other rapidly evolving technologies, such as artificial intelligence (AI), blockchain, and robotics. High immersion, interactivity, and user engagement in VR leverage and compound the organizational potential of these other emerging technologies. For example, AI-supported data visualizations can be brought into VR to help decision-makers steer organizational actions according to different trends and scenarios. The use of digital voice agents (DVAs), such as Google



Virtual reality as a critical, novel technology that can transform how organizations conduct their operations.



Assistant or Microsoft's Cortana, can help users complete different routine tasks in VR. Additionally, blockchain holds potential for fostering secure ownership and transfer of digital assets in VR. 5G networks enable VR to be used as an immersive interface for robotic teleoperations where, for example, the user's body motions can help achieve utmost accuracy.²¹ The possibilities are practically endless; in the future, advancements in brain-computer interfaces (BCIs) provide fascinating possibilities where the use of VR could be, at least partly, controlled by brain signals.^{21,22,33}

VR as a Communication Platform: High-Performing Remote Communication

Every meaningful action in an organization, such as knowledge creation or decision-making, tends to depend on the success of communication and information transfer.⁷ Therefore, the content in VR with the most potential is other people. Implementing communication features even in the simplest use cases, such as a virtual sales meeting in VR, can significantly leverage their potential. Accordingly, when communication features are integrated in more complex use cases, such as industrial design, their potential benefits continue to grow. When VR is used as a communication platform, it can be referred to as social virtual reality (SVR).

Next, we extend our analysis with three reasons why VR is a game-changing communication platform. Specifically, we describe how SVR enables multi-user social interaction that simulates real-life communication and extends it to new forms of remote work.


Reason 4: Every communication process can be simulated

A lack of face-to-face communication deteriorates the richness of communication in organizations. Deriving the most out of current communication tools can mitigate this problem but not fix it. In general, discussions, dialogue, and problem-solving benefit from synchronous communication (for example, video conferencing), whereas the transfer of a large amount of diverse and new information tends to benefit from asynchronous communication (for instance, email).⁹


SVR supports both of these fundamental communication processes—synchronous and asynchronous—in an intuitive and natural manner. Most importantly, SVR can simulate and extend face-to-face communication in a spatial setting. For example, 3D models can be loaded for discussion and dialogue, which fosters users’ shared sense-making and understanding of how others interpret the available information. In contrast, text- or voice-based annotations provide an important feedback mechanism, where users are able to guide, assist, or exchange ideas more elaborately without time constraints. Annotations that are placed directly on 3D objects also maintain the context in communication. Ideally, SVR substitutes many different communication channels by merging them into one. Instead of a plethora of email discussions or video conferencing sessions, every detail from, for example, a product design pipeline, can be discussed and commented on in SVR.

SVR that includes tools for presentations and brainstorming, such as file sharing, whiteboards, and sticky notes, extends a physical meeting room to a virtual sphere. Avatar-based interaction, natural 3D space, and spatial sound enable multiple real-time discussions, where participants interact and communicate spatially as opposed to looking at each other on a monitor. In general, authentic spatial collaboration significantly enhances an individual’s acquisition of professional skills, because it allows them to observe how others behave and operate.⁸ Thus, connecting spatial communication with task-related content can make VR an ideal platform for collaboration and learning. One of the biggest advantages of SVR is also that the context of communication can be filtered to precisely fit the task at hand, excluding any outside distractions³⁵. Due to the sensory immersion provided by HMDs, the task-related focus can be strictly controlled and maintained in SVR.

Theoretically, SVR can facilitate every communication process imaginable and, thus, potentially exceed communication effectiveness compared to real-world settings. For example, one can follow a live keynote presentation, rewind to watch parts of it again, and then catch up with others, just like



Current methods for examining complex information via 2D displays impose obvious limitations on the presentation of information to users.



pressing fast-forward on a television set¹. Additionally, SVR provides communication tools that are not available in the real world, such as a laser pointer coming out directly from an avatar’s fingertip. As another example, avatar profiles as “floating billboards”¹ can disclose a participant’s name, role in the organization, competencies, or other relevant information that we sometimes fail to remember about our colleagues. Perhaps disclosing personal interests in avatar profiles would generate informal social bonding that is otherwise difficult to achieve remotely.

Informal communication is something that organizations struggle to maintain in remote work. It is well known that informality is critical in terms of networking and generating innovations and new ideas. Some top executives are worried that extensive remote work during the COVID-19 pandemic will lead to a decrease in informality.²⁵ SVR provides an especially promising position for tackling this issue with informal virtual spaces, which can be just like a virtual version of a company’s physical break room, characterized by the richness of communication and lack of formal rules, roles, and timetables. Informal virtual spaces can be used anytime, anywhere, without disrupting formal work processes.¹¹ Similarly, SVR can also facilitate social networking and maintaining work-related social relations at virtual events.¹⁰

Reason 5: Transformed group dynamics

Organizational group dynamics, such as trust development, are extremely difficult to manage in conventional remote work.²⁹ However, one of the novelties of avatar-based communication in SVR is its ability to facilitate many fundamental conscious and subconscious social interactions in a spatial setting. Avatar-based communication mimics the sensation of participants being with distant others physically. Just like physical bodies, avatars are both communicative tools and display systems. We communicate via avatars and our behavior allows others to sense and predict our emotions and intentions. Research shows that this behavior is largely automatic.² Today, much of this behavior—posture, interpersonal distance, gaze, and facial move-

ments—can be tracked and displayed in VR, which opens up interesting business possibilities (and data privacy issues) for exploiting the user’s behavioral or even biometric¹⁶ data in VR.²

Of course, current SVR technology is often based on cartoonish avatars that are not yet able to display fully realistic body language or facial expressions. However, even the most basic forms of nonverbal communication, such as the gaze, can have a significant effect on communication performance. For example, the gaze communicates points of interest and, thus, fosters turn-taking and dialogue.¹ Recent advances in VR-related tracking technologies suggest that the avatar gaze, just like realistic avatar hand and facial movements, will soon be a standard feature of SVR.¹⁶ Developments in these tracking technologies are critical because they affect the avatar’s behavioral realism and the user’s nonverbal communication performance.

It is well known that collaboration performance in remote work is built on strong interpersonal trust. However, conventional remote communication tools have raised different trust-building issues due to individuals’ inability to physically and spatially observe how others behave and operate.²⁹ Although SVR does not yet offer fully realistic social simulation, it already holds tremendous potential for enhancing different trust-building mechanisms. As different formal and informal activities are increasingly integrated into SVR, users are able to learn more from others’ skills and personalities and build shared experiences that are comparable to the ones from the physical world. Interestingly, a brain imaging study shows that the trust-building process in avatar-based communication is quite similar to that in face-to-face communication, except that real facial information works better when forming initial trust (that is, trust between strangers or acquaintances).²⁸ There is already commercial interest in building photorealistic avatars for VR, and they are expected to arrive in the coming years.³⁰

Recent studies also suggest that reciprocal communication and behavioral realism seem to mitigate the uncanny valley—the “eerie sensation” users get when viewing almost, but not perfectly, photorealistic artificial

faces.³¹ This development can have interesting implications for the adoption of SVR in a highly formal work context, such as business meetings. But for now, why not satisfy our natural tendency to trust real human faces by embedding video conferencing into SVR?

However, SVR also allows users to display an altered version of themselves by customizing their avatars. Avatar customization is not just a novelty issue or something that connects only with consumer VR and entertainment. It is a powerful tool for nonverbal communication and online identity management. Studies show that avatar characteristics may have psychological and behavioral implications—a phenomenon known as the Proteus effect.³⁶ For example, Yee et al³⁶ show that taller avatars performed better in a negotiation task and attractive avatars disclosed more personal information. Further, the avatar’s nonverbal behavior can be modified, filtered, or automated to not display the user’s actual nonverbal behavior.¹ For example, an artificial smile (that is, an avatar’s smile that is enhanced with algorithms) can leave everyone in a better mood after a virtual conferencing session.²⁶ How the Proteus effect and nonverbal modifications can transform group dynamics and information transfer in SVR holds much promise for future remote work.

Reason 6: AI agents as organizational actors

A vast amount of relevant information gets lost in organizational communication due to our limited information-processing capabilities.⁷ However, introducing AI avatars—or agents—into SVR allows completely new forms of collaboration and information-sharing practices for organizations. Technology’s “human-likeness” can affect how individuals interact with and form attitudes toward technology. Thus, if a technological entity looks and acts like a human, it is more likely to be perceived as, for example, “competent” instead of “functional.”²⁰

A variety of AI capabilities that mimic the human mind (for example, reasoning, object and speech recognition, and a dialogue system) can be attached to agents, and these capabilities can be expanded further with, for example, big data analytics.^{14,24} Especially in

knowledge-intensive work, agents can take an interesting position in different knowledge-creation and decision-making activities when an organization’s stakeholders interact with each other and agents. In SVR, interactive agents can be available at all times, and their communication and information-sharing capabilities will increase in parallel with different AI developments.

Of course, agents could conduct different routine or assistive tasks comparable with the use of current chat bots or DVAs. However, unlike conventional AI, agents are also perceived as physical entities. For example, agents can physically navigate users through a virtual event or illustrate how to perform various organizational tasks, such as machine maintenance. Some training activities, for example, could be scripted using activities performed by real human users, tackling some issues with scalable content creation. Furthermore, one especially interesting domain for agents is sales and marketing. Agents can represent organizations in the digital realm in a scalable manner. Even an agent’s nonverbals can soon be simulated according to the potential client’s cultural background and preferences. If needed, a human user can be summoned to replace the AI. For example, when a customer wants more detailed information in a sales situation, the right salesperson with proper language preferences can take control of the AI’s avatar.

The potential of agents as organizational actors probably increases with their behavioral realism. Some scholars describe a future where agents display increasingly human-like behavior, such as being able to mimic our nonverbal cues and emotions.^{22,31} For example, agents might be able to detect our emotions from our voice pitch and facial information (our facial movements can already be tracked in VR). Agents could also create believable reciprocal communication patterns, and communication with agents could, thus, become nearly indistinguishable from human-to-human communication.²² As a practical example, see the Seymour et al³¹ study that presents Baby X, a computer-controlled agent.

Conclusion and Implications

VR is finally reaching a point in its development where it can be widely used

Table 1. VR as a computing platform—key implications for organizations.

Key aspect of VR	Potential organizational benefits	Key actions for realizing the benefits of VR
1. Enriched data and information	<ul style="list-style-type: none"> ▶ Enhanced organizational knowledge creation and decision-making ▶ Reduced misunderstandings and uncertainty ▶ Increased stakeholder engagement ▶ Enhanced stakeholder understanding and recall of complex or domain-specific information 	<p>Content creation for VR</p> <ul style="list-style-type: none"> ▶ Discover existing and novel forms of digital assets that could benefit from being viewed, stored, organized, and retrieved in VR ▶ Adapt content in VR according to stakeholder needs and preferences <p>Capacity-building for VR</p> <ul style="list-style-type: none"> ▶ Map out and create awareness for an organization's stakeholders who could benefit from the use of VR ▶ Develop capabilities for novel knowledge-management practices required in VR
2. Immersive workflows and training	<ul style="list-style-type: none"> ▶ Workflows and training with unrestricted participation and interactions ▶ Enhanced acquisition of professional skills and knowledge ▶ Highly iterative and effective collaboration and knowledge transfer ▶ Enhanced user-flow experience and motivation to perform tasks efficiently 	<p>Workflow creation for VR</p> <ul style="list-style-type: none"> ▶ Discover existing and novel workflows that benefit from VR-enriched data and information and have a sense of natural interactions in a spatial setting ▶ Enable persistent content for iterative workflows and project management <p>Implementation of training and simulations in VR</p> <ul style="list-style-type: none"> ▶ Prioritize training or simulation scenarios that would be impossible, dangerous, or costly to perform in real life ▶ Introduce VR in both hard- and soft-skills training ▶ Enrich training and simulation scenarios with playful and gamified elements
3. Increasing synergies with other IS and emerging technologies	<ul style="list-style-type: none"> ▶ Fluent information transfer between different IS and the organization's stakeholders ▶ Compounding the benefits of various emerging technologies by leveraging the immersive and interactive nature of VR ▶ Novel business opportunities and use cases when VR is integrated with other emerging technologies 	<p>Integration of IS with VR</p> <ul style="list-style-type: none"> ▶ Enable essential information flows about the organization and users between existing organizational IS and VR ▶ Enrich real-time organizational data and information via VR <p>Integration of emerging technologies with VR</p> <ul style="list-style-type: none"> ▶ Identify synergies between emerging technologies and organizational data, information, and workflows ▶ Incrementally introduce VR solutions that exploit the technological development of emerging technologies

Table 2. VR as a communication platform—key implications for organizations.

Key aspect of VR	Potential organizational benefits	Key actions for realizing the benefits of VR
4. Every communication process can be simulated	<ul style="list-style-type: none"> ▶ High-performing remote communication and collaboration ▶ Enhanced dialogue and shared understanding ▶ Enhanced transfer of context-bound data and information ▶ Possibility to control the task-related focus ▶ Remote, casual interactions and networking 	<p>Facilitation of interpersonal communication in VR</p> <ul style="list-style-type: none"> ▶ Introduce real-time, avatar-based interaction ▶ Enable the use of customized avatar profiles with individualizing information ▶ Enable new ways of avatar-based interaction that are not possible in the real world <p>Facilitation of formal and informal meetings and events in VR</p> <ul style="list-style-type: none"> ▶ Integrate task-related communication tools into VR environment ▶ Filter the context of communication according to tasks ▶ Exploit the use of avatar profiles in networking ▶ Build content and interactions for informal bonding
5. Transformed group dynamics	<ul style="list-style-type: none"> ▶ New forms of online group dynamics and social bonding ▶ Novel trust-building mechanisms in a shared spatial setting ▶ Enhanced online identity management with potential behavioral implications 	<p>Creation of realistic avatars in VR</p> <ul style="list-style-type: none"> ▶ Increase avatars' behavioral realism via advanced tracking technologies, such as eye, face, and body tracking ▶ Increase avatars' photorealism for communication processes that are highly formal or emphasize trust-building between unacquainted individuals <p>Introduction of nonverbal avatar enhancements in VR</p> <ul style="list-style-type: none"> ▶ Enable rich avatar customization as an online identity management system and to exploit the Proteus effect ▶ Build algorithms that modify, filter, or automate a user's nonverbal expressions and gestures in VR
6. AI agents as organizational actors	<ul style="list-style-type: none"> ▶ Novel remote collaboration and knowledge-creation practices ▶ Agent-supported training and tutoring ▶ Agents as scalable organizational actors and physical entities in the digital realm 	<p>Creation of agents to support knowledge-intensive work in VR</p> <ul style="list-style-type: none"> ▶ Build agents that provide information and support in repetitive or routine tasks ▶ Build agents with advanced AI capabilities that provide support in problem-solving and decision-making activities <p>Creation of agents as physical entities in VR</p> <ul style="list-style-type: none"> ▶ Build reciprocal nonverbal communication patterns for agents ▶ Enable the control and learning of practical and task-related skills for agents

to support and enhance various work tasks in organizations. However, its uniqueness as a computing and communication platform is still not widely understood. Our article builds upon VR's well-known potential to foster human cognitive functions in simulated environments and specifically aims at shedding light on its organizational implications in the context of knowledge management and remote communication. Based on a review of scientific literature and practice-driven insights, we have outlined six reasons why VR is a game-changing technology for organizations. As a computing platform, VR enables novel knowledge-management practices for managing enriched data and information and immersive workflows, which both benefit greatly from integrations with appropriate IS and other emerging technologies, such as AI. As a communication platform, VR can simulate every communication process imaginable (some of which can be AI-supported), which has significant potential for fostering an organization's online communication performance, knowledge creation, and group dynamics.

One of the main takeaways of this article is that VR enables not just substituting the physical with virtual but also novel ways of working. VR can make existing work more effective, but it can also bring completely new business opportunities for organizations. We elaborate these potential benefits for organizations in Table 1 and Table 2. Due to rapid developments in VR technology, organizations have not yet exploited these various possibilities afforded by the newest VR hardware and software. It is important for organizations to identify the business processes where the easily capturable benefits of VR converge with ease of adoption. As with any new innovation, organizations will need to develop new skills and capabilities to export their relevant digital assets, interactions, and communication processes to VR.

With sufficient capabilities, VR can also be used to radically transform organizational operations. However, VR is not a one-size-fits-all solution. Its benefits often emerge in very specific use cases (such as a particular simulation) that do not necessarily translate to a

monetizable VR service that could serve a larger group of companies. Instead, VR development is often based on customized solutions, which has made it difficult to scale and adapt them to different organizational contexts.

This article has identified the benefits of VR specifically for the context of knowledge management and remote communication in order to obtain key insights about the game-changing nature of VR for organizations. We also provide several key actions in Tables 1 and 2 that organizations can carry out to take full advantage of the six key aspects of VR we described and to realize the organizational benefits thereof.

References


1. Bailenson, J., Beall, A., Loomis, J., Blascovich, J., and Turk, M. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments* 13, 4 (2004), 428–441.
2. Bailenson, J. Protecting nonverbal data tracked in virtual reality. *JAMA Pediatrics* 172, 10 (2018), 905–906.
3. Bailenson, J. Is VR the future of corporate training? *Harvard Business Review* (Sept. 18, 2020), <https://hbr.org/amp/2020/09/is-vr-the-future-of-corporate-training>
4. Benford, S., Greenhalgh, C., Rodden, T., and Pycocock, J. Collaborative virtual environments. *Communications of the ACM* 44, 7 (2001), 79–85.
5. Bryson, S. Approaches to the successful design and implementation of VR applications. In R. Earnshaw, J. Vince, and H. Jones, eds., *Virtual Reality Applications*. San Diego, CA, Academic Press (1995), 3–15.
6. Bryson, S. Virtual reality in scientific visualization. *Communications of the ACM* 39, 5 (1996), 62–71.
7. Choo, C.W. The knowing organization: How organizations use information to construct meaning, create knowledge, and make decisions. *Inter. J. Information Management* 16, 5, (1996), 329–340.
8. Dede, C. Immersive interfaces for engagement and learning. *Science* 323 (2009), 66–69.
9. Dennis, A.R., Fuller, R.M., and Valacich, J.S. Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly* 32, 3 (2008), 575–600.
10. Educators in VR (2020), <https://educatorsinvr.com/events/>
11. Fairs, M. 'Incredible things are happening' in virtual reality, say architects in lockdown. *Dezeen* (May 13, 2020). <https://www.dezeen.com/2020/05/13/incredible-virtual-reality-coronavirus/>
12. Faughnder, R. 'The Lion King's' VR helped make a hit. It could also change movie making. *LA Times* (July 26, 2019). <https://www.latimes.com/entertainment-arts/business/story/2019-07-26/disneys-lion-king-remake-is-a-hit-its-virtual-reality-technology-could-change-how-movies-are-made>
13. Gorlick, A. The productivity pitfalls of working from home in the age of COVID-19. *Stanford News* (Mar. 30, 2020). <https://news.stanford.edu/2020/03/30/productivity-pitfalls-working-home-age-covid-19/>
14. Gupta, S., Kar, A., Baabdullah, A., and Al-Khowaiter, W.A. Big data with cognitive computing: A review for the future. *Intern. J. of Information Management* 42 (2018), 78–89.
15. Horwitz, J. Boeing will use Varjo VR to train astronauts for Starliner missions. *VentureBeat* (June 11, 2020). <https://venturebeat.com/2020/06/11/boeing-will-use-varjo-vr-to-train-astronauts-for-starliner-missions/>
16. Horwitz, J. HP's Reverb G2 Omnicept VR headset adds heart, eye, and face tracking. *VentureBeat* (Sept. 30, 2020). <https://venturebeat.com/2020/09/30/hps-reverb-g2-omnicept-vr-headset-adds-heart-eye-and-face-tracking/>
17. Jalo, H., Pirkkalainen, H., Torro, O., Lounakoski, M., and Puhto, J. Enabling factors of social virtual reality

- diffusion in organizations. In *Proc. of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference*, (2020).
18. Kohler, T., Fueller, J., Matzler, K., and Stieger, D. Co-creation in virtual worlds: The design of the user experience. *MIS Quarterly* 35, 3 (2011), 773–788.
19. Krokos, E., Plaisant, C., and Varshney, A. Virtual memory palaces: Immersion aids recall. *Virtual Reality* 23, 1 (2019), 1–15.
20. Lankton, N.K., Mcknight, D.H., and Tripp, J. Technology, humanness, and trust: Rethinking trust in technology. *J. of the Assoc. for Information Technology* 16, 10 (2015), 880–918.
21. LaValle, S. *Virtual Reality*. Cambridge University Press (2020).
22. Metzinger, T.K. Why is virtual reality interesting for philosophers? *Frontiers in Robotics and AI* 5 (2018), 101.
23. Mindesk: Real-time VR CAD. (2020). <https://mindeskvr.com/>
24. Modha, D.S., Ananthanarayanan, R., Esser, S.K., Ndirango, A., Sherbondy, A.J., and Singh, R. Cognitive computing. *Communications of the ACM* 54, 8 (2011), 62–71.
25. What Satya Nadella thinks. *New York Times DealBook* (May 14, 2020). <https://www.nytimes.com/2020/05/14/business/dealbook/satya-nadella-microsoft.html>
26. Oh, S.Y., Bailenson, J., Krämer, N., and Li, B. Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PIOS One* 11, 9 (2016), e0161794.
27. Paes, D., Arantes, E., and Irizarry, J. Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems. *Automation in Construction* 84, (2017), 292–303.
28. Riedl, R., Mohr, P.N., Kenning, P.H., Davis, F.D., and Heekeren, H.R. Trusting humans and avatars: A brain imaging study based on evolution theory. *J. of Management Information Systems* 30, 4 (2014), 83–114.
29. Robert, L.P., Denis, A.R., and Hung, Y.-T.C. Individual swift trust and knowledge-based trust in face-to-face and virtual team members. *J. of Management Information Systems* 26, 2 (2009), 241–279.
30. Rubin, P. Facebook can make VR avatars look—and move—exactly like you. *Wired* (Mar. 13, 2019). <https://www.wired.com/story/facebook-oculus-codec-avatars-vr/>
31. Seymour, M., Riemer, K., and Kay, J. Actors, avatars and agents: Potentials and implications of natural face technology for the creation of realistic visual presence. *J. of the Assoc. for Information Systems* 19, 10 (2018), 953–981.
32. Slater, M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557.
33. Slater, M. and Sanchez-Vives, M.V. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI* 3, 74 (2016).
34. Srivastava, S.C. and Chandra, S. Social presence in virtual world collaboration: An uncertainty reduction perspective using a mixed methods approach. *MIS Quarterly* 42, 3 (2018), 779–804.
35. Steffen, J.H., Gaskin, J.E., Meservy, T.O., Jenkins, J.L., and Wolman, I. Framework of affordances for virtual reality and augmented reality. *J. of Management Information Systems* 36, 3 (2019), 683–729.
36. Yee, N., Bailenson, J.N., and Ducheneaut, N. The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research* 36, 2 (2009), 285–312.
37. Yoon, T.E. and George, J.F. Why aren't organizations adopting virtual worlds? *Computers in Human Behavior* 29, 3 (2013), 772–790.
38. Zuckerberg, M. Facebook's online question-and-answer session. *Facebook* (June 30, 2015). <https://www.facebook.com/zuck/posts/10102213601037571>

Osku Torro is a doctoral researcher at Tampere University, Finland.

Henri Jalo is a doctoral researcher at Tampere University, Finland.

Henri Pirkkalainen is an associate professor of information and knowledge management at Tampere University, Finland.

 This work is licensed under a <http://creativecommons.org/licenses/by/4.0/>

DOI:10.1145/3474359

Experienced and aspiring computing professionals need to manage their qualifications according to current market needs. That includes certification achievement as well as formal education, experience, and licenses.

BY MARK TANNIAN AND WILLIE COSTON

The Role of Professional Certifications in Computer Occupations

THE COVID-19 PANDEMIC has disrupted employment and education in the U.S. and other countries. Experienced and aspiring practitioners within computing occupations need to manage their qualifications with respect to current market needs. Achievement of professional certifications along with formal education, experience, and licenses provides the basis for employment qualifications.

This article focuses on certifications in the context of the U.S. job market across a range of computer occupations, such as software developer, information security analyst, computer network architect, and Web

developer, as described by the U.S. Bureau of Labor Statistics (BLS). Practitioners working outside of the U.S. may find that the market analysis discussed here differs from their own experiences. Analysis of the U.S. market provides context for the discussion of certification attributes to consider when selecting a certification for investment of one's time, reputation, and money.

This article aims to help answer the following questions:

- ▶ How does certification affect employment?
- ▶ How does one pick a certification?
- ▶ What are the consequences of any particular choice?

While comprehensive answers to these questions cannot be provided in one article, this article should help frame answers to them.

This article investigates a number of dimensions related to certifications and focuses on the following (in order):

- ▶ Certification demand
- ▶ Motivations
- ▶ Value
- ▶ Quality
- ▶ Cost

Certification Demand

To begin exploring the relationship between certification and employment in the U.S., certification demand is viewed from two perspectives. The first is from

» key insights

- **U.S. certification demand varies across computer occupations. Most Information Security listings show an interest in a certification while almost none of the Computer and Information Research Scientist listings do.**
- **Beyond employment, certifications can help to develop proficiency in topic areas not readily accessible at work.**
- **Certification value is largely outside the certification holder's control. Those who one encounters will assess its value.**
- **While value is often based on certification quality, a high-quality, lesser-known certification is likely to be undervalued.**
- **Beyond attainment costs, consider a particular certification's long-term maintenance and opportunity costs.**



an employee's perspective as observed in the BLS Current Population Survey (CPS). The second is the employer's perspective, which is based on analysis of requested professional certifications found in job listings across the U.S.

Employee's Perspective

CPS is a monthly survey based on U.S. household responses as opposed to a survey of businesses. This survey includes questions regarding work-related licensing and certifications. Unlike certifications, licenses are issued by a government entity and are required in order to practice legally specified occupations.

In 2019, Cunningham³ reported on results related to these questions, and all values cited from this report are annual 2018 averages. Cunningham reports an average of 13.5% of U.S. workers within the Computer and Mathematical

occupations state they have a license or certification.³ Of that 13.5%, 6.6% have at least one certification. Cunningham reports that those with associate degrees within Computer and Mathematical occupations are the largest group by formal education to hold certifications or licenses (19%). Of the workers holding master's degrees, 14.5% have a certification, and doctoral degree holders slightly edge out bachelor's degree holders, with 13.3% versus 13.0%, respectively. The CPS shows that 413,000 (59.9%) employees in Computer and Mathematical occupations were required, according to their positions, to attain a license or certification while 277,000 (40.1%) reported that their certifications or licenses were not required.³

Based on these statistics, the report indicates a minority of Computer and Mathematical professionals has a license or at least one certification. Of

those who do, a majority of practitioners is required to have them. Computer and Mathematical practitioner responses show licensing or certifications are useful for completing or enhancing qualification portfolios regardless of the level of formal education achieved.

Employer's Perspective

Background. Before presenting employer certification-demand findings, it is necessary to describe the methodology in order to assist in interpreting the results. The certification-demand analysis was performed using the Economic Modeling Specialists International (EMSI) dataset. To populate the dataset, EMSI combs through 100,000 websites, effectively capturing job listings for more than 1.5 million companies. The same job listings regularly appear on multiple websites. To reduce duplicates, EMSI uses a machine learn-

ing-based duplicate-detection process.

Remote or location-neutral listings (for example, a search for a single team lead who can be located in either Atlanta, New York, or Los Angeles) are counted as unique listings from within each targeted job market, despite referencing the same opening. An extract from the EMSI dataset for March 2019 through February 2020 was constructed for the BLS Standard Occupation

Classification (SOC) codes consistent with the Computer and Information Technology occupations (15-1100) within the Computer and Mathematical category (15-0000). There are 12 distinct subcategories under 15-1100 and one general catchall category. EMSI uses the BLS classification scheme to organize the hundreds of thousands of listings it collates.

There are two major sources of am-

biguity in these results. The first is that an opening that requests one certification from a list of alternative certifications is counted as many times as the number of certification alternatives listed within the job listing. Job listing over-counting is most prevalent in occupation categories where there are a variety of desirable alternatives.

The second source of ambiguity is that not all certification entries represent a distinct certification. Some reference families or categories of certifications. For instance, Global Information Assurance Certifications (GIACs) are a family of certifications issued by the SANS Institute. Some job listings, however, specify particular certifications within the family, such as GIAC Security Essentials Certification. For this article, references to categories or families of certifications were left intact to preserve the more general tenor of an employer's interest.

Results. Top-15 analysis was used on records from 10 job subcategories, which is 77% of Computer occupations by BLS subcategory count. This article presents detailed results for three subcategories: information security analysts, computer network architects, and software developers. Top-15 certification analysis counts and ranks certification name appearance, which helps in understanding the popularity of or interest in particular certifications. An important attribute of this analysis is the number of records not represented by the set of identified top-15 certifications. This quality of representation is called certification-demand coherence.

The Information Security Analysts (15-1122) subcategory in Table 1 had the highest level of certification-demand coherence, with top-15 certifications being requested in 95% of more than 345,000 listings. Computer Network Architects (15-1143), in Table 2, was the second highest, with top-15 certifications being requested in 52% of more than 38,000 listings. Table 3 focuses on the BLS Software Developers (15-1132) subcategory. EMSI data shows that only 2.72% of more than 1,780,000 listings are represented by its particular top-15 subset of certifications.

The overall trend of certification-demand coherence across occupation subcategories rapidly drops beyond

Table 1. Top 15 requested certifications for Information Security Analysts (15-1122).

Certifications	Listings	Representation
Total	345,207	100.0%
<i>Certified Information Systems Security Professional</i>	90,496	26.2%
<i>GIAC Certifications</i>	41,508	12.0%
<i>Certified Information Security Manager</i>	32,150	9.3%
<i>Certified Information System Auditor (CISA)</i>	31,701	9.2%
<i>CompTIA Security+</i>	26,804	7.8%
<i>Certified Ethical Hacker</i>	23,372	6.8%
<i>GIAC Certified Incident Handler</i>	12,918	3.7%
<i>GIAC Security Essentials Certification</i>	10,837	3.1%
<i>IAT Level II Certification</i>	10,767	3.1%
<i>Cisco Certified Network Associate</i>	10,615	3.1%
<i>Certified In Risk and Information Systems Control</i>	7,959	2.3%
<i>Offensive Security Certified Professional</i>	7,549	2.2%
<i>NIST Cybersecurity Framework (CSF)</i>	7,353	2.1%
<i>Systems Security Certified Practitioner</i>	7,327	2.1%
<i>Cisco Certified Security Professional</i>	6,724	1.9%
Unrepresented Listings	17,127	5.0%

Note: Italicized entry indicates a certification's main focus aligns with occupation's unique responsibilities.

Table 2. Top 15 requested certifications for Computer Network Architects (15-1143).

Certifications	Listings	Representation
Total	38,515	100.0%
<i>Cisco Certified Network Professional</i>	4,795	12.4%
<i>Cisco Certified Network Associate</i>	4,321	11.2%
<i>Cisco Certified Internetwork Expert</i>	3,848	10.0%
<i>CompTIA Security+</i>	1,091	2.8%
<i>Certified Information Systems Security Professional</i>	803	2.1%
<i>Cisco Certified Design Professional</i>	668	1.7%
<i>Juniper Networks Certified Internet Expert</i>	597	1.6%
<i>IAT Level II Certification</i>	588	1.5%
<i>ITIL Certifications</i>	586	1.5%
<i>Project Management Professional Certification</i>	548	1.4%
<i>Microsoft Certified Systems Engineer</i>	536	1.4%
<i>Juniper Network Certified Internet Professional (JNCIP)</i>	492	1.3%
<i>Juniper Networks Certified Internet Associate</i>	440	1.1%
<i>CompTIA Network+</i>	368	1.0%
<i>ITIL Foundation Certification</i>	364	0.9%
Unrepresented Listings	18,470	48.0%

Note: Italicized entry indicates a certification's main focus aligns with occupation's unique responsibilities.

the first three occupation subcategories as seen in Figure 1. The likely reasons for a listing not to be represented are either a certification is listed but not present among the top 15 or no certification is mentioned within the listing. These reasons act as opposite ends of a continuum on which the overall unrepresented listing collection falls. To better understand the first reason, the certification market for a particular occupation subcategory may have many desirable certifications, thus spreading out demand beyond 15 certifications. A possible example is subcategory 15-1151, which has 17.9% demand coherence and 12 out of 15 certifications with occupational focus. To explain the second reason, it is possible that an employer is unaware of suitable available certifications or disinterested in requesting a certification for a particular listing.

The remaining occupation areas are described at a summary level in Table 4. Nearly four million listings did not request any of the certifications identified in any of the top-15 analyses performed across the 10 subcategories. Given the idiosyncrasies in job-listing counting and certification labeling, these results align fairly well with Cunningham’s finding of 6.6% certification achievement reported in the CPS. The last six occupation areas listed in Table 4 appear to indicate that certifications are of low interest in computing occupations. When considering the first four occupation areas, practitioners would be wise to seek out at least one certification. Considering that diversified certification demand within an occupation area undermines the ability of top-15 analysis to gauge overall certification demand, further investigation into appealing job opportunities is recommended prior to dismissing the need for certifications in general.

Motivations

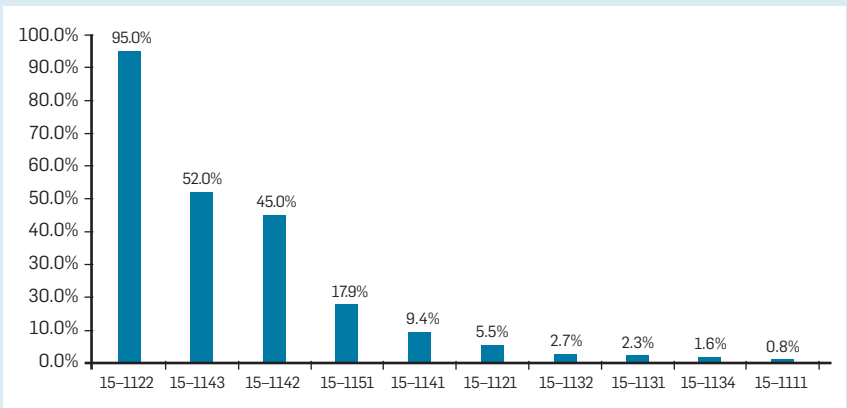
There are a number of stakeholders who have a vested interest in certifications, such as candidates/employees, employers, and issuing authorities. Each has a distinct rationale for their interest in or involvement with certifications. These motivations contribute to the dynamics within the approximately \$US2.6 billion worldwide training and certification marketplace.¹⁴ This sec-

Table 3. Top 15 requested certifications for Software Developers (15-1132).

Certifications	Listings	Representation
Total	1,789,139	100.0%
<i>Salesforce Certification</i>	5,827	0.3%
Microsoft Certified Professional	5,193	0.3%
ITIL Certifications	4,620	0.3%
Microsoft Certified Systems Engineer	4,493	0.3%
Project Management Professional Certification	3,335	0.2%
Salesforce Certified Administrator	3,038	0.2%
IAT Level II Certification	3,006	0.2%
<i>Certified Scrum Master</i>	2,927	0.2%
GIAC Certifications	2,917	0.2%
ITIL Foundation Certification	2,722	0.2%
Cisco Certified Network Associate	2,484	0.1%
Certified Power Quality Professional	2,480	0.1%
Microsoft Certified Systems Administrator (MCSA)	2,264	0.1%
<i>AWS Certified Solutions Architect</i>	1,776	0.1%
VMware Certified Professional (VCP)	1,559	0.1%
Unrepresented Listings	1,740,498	97.3%

Note: Italicized entry indicates a certification’s main focus aligns with occupation’s unique responsibilities.

Figure 1. Certification-demand coherence by BLS SOC code. SOC codes are explained on the BLS website.²



tion explores motivations among candidates/employees and employers.

Candidates or Employees. Although certification-demand coherence is currently soft for seven of the 10 categories in the U.S., certifications should be anticipated to be either an explicit or implicit requirement for positions in the first four occupation areas represented in Figure 1. Labor supply and hiring practices within a geographic region relevant to an open position dictate how flexible posted certification requirements are.

The demand for certification is more pronounced in a number of non-U.S. job markets, such as the U.K., the

EU, Canada, Pakistan, India, Singapore, Thailand, and Japan, where a credible certification confers standing within an area of competency.^{8,17} Often, an arbitrary certification does not satisfy or exceed expectations. A candidate should consider an employer’s expressed certification requirements for opportunities in job subcategories with low demand coherence.

Employees may be challenged through their performance review process to pursue professional development by achieving or maintaining employer-desired certifications. As seen in the BLS CPS discussion, roughly 40% of certification holders

are not required to have certifications. One motivation for these CPS respondents is that their employers look favorably on an employee’s commitment to competency development and may award greater compensation to preserve their value in-house. Independent of employment and compensation considerations, achieving certifications can lead to personal growth and confidence.

Working and remaining relevant in dynamic computing occupations requires practitioners to seek lifelong learning. Kruchten raises the notion that relevancy of knowledge and skills in software development has a half-life, which is a notion applicable to many technology fields.¹⁰ Certifications facilitate lifelong learning by providing a well-defined, compiled set of knowledge and skills, often supported by accessible learning resources—for instance books, courses, videos—and an assessment to gauge proficiency. Continued learning after initial certification is thought to be as valuable as the initial accomplishment because most computing certification knowledge areas regularly experience change.

At work, opportunities may not track with market changes. A current employer may be unable or unwilling to pursue certain prevalent technologies. When it comes time to leave, the mismatch between the current state of the market and actual experience may present challenges when seeking new work. As candidates, practitioners may find relevant certifications as an adequate means to compensate for experiences that do not track well with popular practices and technology require-

ments. Although ambiguity regarding certifications in a job listing raises concern, close certification-topic alignment with a position’s responsibilities is more likely to be taken into positive consideration.

Employers. The working knowledge and available skillset of each employee, brought to bear each day, has the potential to help the organization realize its strategy, generate revenue, or fulfill its mission. With each organization being unique, training new hires in the essential policies, processes, and systems cannot be avoided. If the labor supply is robust with relevant qualifications, employers are able to limit the amount of in-house training for role-relevant knowledge and skills. In addition to specifying the types and levels of formal education, certifications can be a convenient means of specifying the types of knowledge and skills needed for a particular role.

Competition in a vigorous market urges a business to produce better products or deliver more desirable services at a sustainable price before the competition. This often hinges on the development and adoption of new technology—built in-house or provided by other sources. Designing, implementing, and maintaining new technology requires people who can work through technical complexities and unknowns. The ability to diverge from an established strategic direction to another requires foresight, an agile workforce, and the enablement of business processes. Businesses that are motivated to stay competitive require a risk-taking workforce that is constantly learning. Encouraging em-

ployees to learn new technologies and their real-world application improves the workforce’s readiness to pivot to a new strategic path.

Certifications on new topics are often available sooner than new course content within postsecondary education programs. Technology providers that are motivated in seeing their technology being used provide training and certification soon after product availability. The \$US2.6 billion market is large and competitive. Aggressive issuing and training organizations produce certification materials for topics they believe eager technologists will embrace. Recognition is critical to a certification’s success, so early market penetration of a credible certification improves its legitimacy, network effect, and reputation as demand for the subject area grows. For example, the Certified Information Systems Security Professional, ranked first in Table 1, first launched around 1994, when commercial Internet was experiencing initial business adoption.

There are risks that the workforce can both precipitate and mitigate. Seemingly mundane decisions and actions may result in life-threatening operational or product failure, malware infestation, intellectual property disclosure, or a customer data breach. In regulated industries, these incidents may result in fines or loss of license to operate. In lieu of licensing for most of the computer occupations discussed here, certifications are the available mechanism to independently verify competence. This may be a motivation for the three security certification entries in Table 2 and two security certification entries in Table 3.

The U.S. Department of Defense (DoD) departmental directive, DoD 85701.01-M, requires verified competence in order for information security controls to be properly managed and for qualified staff to make risk-management decisions.⁴ Banks subject to the U.S. Gramm-Leach-Bliley Act (GLBA) Safeguards Rule are evaluated for the qualifications of the personnel enabling the bank’s comprehensive information security program.⁶ Security is inherently a risk management subject; however, competence in network and system design, software engineering,^{12,13} and other technical knowledge

Table 4. Summary results table by SOC in certification-demand coherence order. SOC codes are explained on the BLS website.²

BLS SOC	Demand Coherence	Focus Aligned Certifications	Total Listings	Listings within Top 15	Listings not within Top 15
15-1122	95.0%	14	345,207	328,080	17,127
15-1143	52.0%	8	38,515	20,045	18,470
15-1142	45.0%	10	558,134	251,293	306,841
15-1151	17.9%	12	654,844	117,063	537,781
15-1141	9.4%	3	109,897	10,301	99,596
15-1121	5.5%	8	547,618	30,232	517,386
15-1132	2.7%	3	1,789,139	48,641	1,740,498
15-1131	2.3%	3	181,105	4,114	176,991
15-1134	1.6%	3	477,820	7,831	469,989
15-1111	0.8%	0	90,259	762	89,497

areas are also necessary for employers to avoid computing-related health, safety, and security risks.

Value

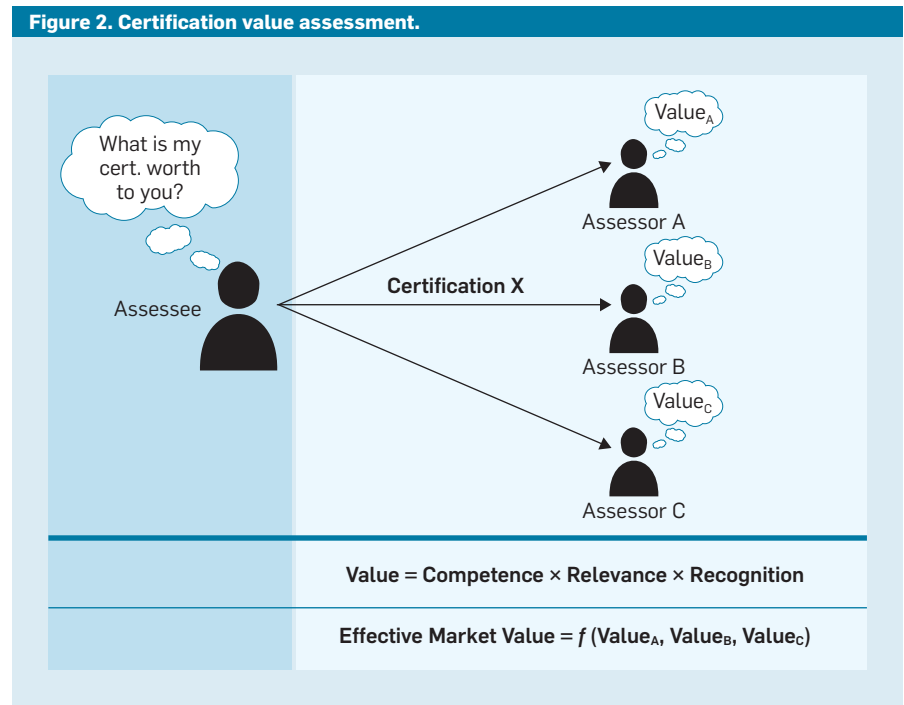
The valuation of a certification involves two parties. The first is the assessee, a stakeholder invested in a particular certification (for example, the issuing authority, certification holder, etc.). The second party is the assessor, who determines if the certification is of value to them (for example, an uncommitted practitioner, employer). As depicted in Figure 2, the value placed on a particular certification depends on the assessor's determination of value. The assessee has committed to the certification and is bound by this value judgment until an opportunity for change is presented. This section explores evaluation of certifications from the assessor's lens.

Three primary attributes of an assessor's determination of value are competence, relevance, and recognition. The product of these three attributes determines value. Competence is the level of subject-area proficiency achieved by a candidate who successfully attains a certification. Relevance relates to a certification's particular fitness-for-purpose. Recognition relates to how easily an assessor can recall the existence and significance of a certification.

For example, a software manager recognizes the significance of CPA as certified public accountant. However, having a CPA is not particularly relevant to filling a software architect role for a customer-service mobile app. In this scenario, the poor fitness-for-purpose terminates value evaluation. The level of accounting competence the CPA represents is disregarded. The effective value of the CPA is zero for this scenario. For practical application of this informal value calculus, the assessor must situate the evaluation of value into a context of need.

Perception of competence and actual competence are often misaligned. Obtaining a certification rarely results in actual competence for a particular certification holder. Select experience, in addition to passing exams, is required for some of the certifications listed in Table 1. Tripp indirectly supports this notion by his caution that the IEEE-CS Certified Software Develop-

Figure 2. Certification value assessment.



ment Professional (CSDP) is not a guarantee of competence. Instead, it is but a measure of an individual's understanding of a certain level of professional practice.¹⁵ At best, a certification is a convincing positive indicator of actual competence. Actual competence is determined in part by experience, personality, intellect, and the sum of a person's entire learning.

Relevance is rarely independent of need. A manager, for instance, evaluates the degree of alignment between the intended role and a certification's associated knowledge, skills, code of ethics, and prerequisites. Managers are unlikely following a formal evaluation rubric. Instead, they often rely on impressions formed by talking to peers, past employee performance, marketing, and possibly performing a task-technology fit analysis.

Knowles et al. investigate cybersecurity certification assessment approaches and their ability to assess competence.⁹ Although Knowles et al. are interested in determining competence assessment usefulness, the assessment approach relates to relevance as well. Consider the Electronics Technicians Association (ETA) International's certifications—some are purely conceptual while others are blends of skills and concepts. An employer looking to fill a senior technical role may find ETA International's hands-on certifications more

relevant than purely conceptual ones. Although difficulty in obtaining a certification often comes to mind with respect to value, difficulty contributes to value if the challenge improves relevance of the certification to a role's responsibilities.

Recognition by an assessor is essential in the value calculus. A well-recognized certification has the potential to impart significant credibility to the assessee if the assessor has an overall positive perception of relevance. Without recognition, actual competence and relevance represented by a certification is unlikely to be acknowledged. Moreover, the challenges the assessee must overcome to obtain the certification are likely to be underappreciated. The value of an unrecognized certification is essentially zero from a particular assessor's perspective. If this is generally true for a given market, the effective value of the certification in this market is near zero. A certification's network effect and branding are essential for assessors to better align their perception of value with actual value.

Although the focus on recognition has been on a specific certification, the intended market needs to recognize the role and potential of certifications in general. Low certification-demand coherence and the appearance of 12 non-development-related certifications in Table 3 appear to indicate low recog-

dition or appreciation for software development-oriented certifications. This observation is consistent with Seidman's analysis of software engineering-certification interest in his 2014 ACM *Inroads* article.¹³ As a practical example, this broad issue may pose a recognition problem for candidates holding the IEEE-CS Professional Software Engineering Master (PSEM) certification, which replaced the aforementioned IEEE-CS CSDP.

Quality

Certification quality drives the competence and relevance value attributes determined by assessors. Given the influence of recognition, low value does not necessarily mean low quality. Quality of a certification results from addressing three interrelated aspects of organizational quality. The first is foundational and relates to how an issuing authority goes about the general business of certification. The second aspect addresses the certification's competency model and how competence is assessed. The final aspect is the body of knowledge and professional behavior specified as the curriculum or the content scope of a certification. The first aspect strongly influences the outcome quality of the other two. Dependence between the second and third aspects exists, because the competency model cannot be fully designed without the content scope specified. Multiple credible certifications can focus on the same competency content but seek to assess different levels of proficiency—for example by evaluating Novice, Competent, or Expert within the Dreyfus competence model.⁵

The relationship between an issuing authority's governance and certification quality is analogous to the relationship between universities' academic administration and the degrees they confer. Accreditation of the university and degree program levels is often a requirement for discerning stakeholders. Along the lines of accreditation of issuing authorities, an international standard, ISO/IEC 17024:2012, has been established. ANSI accredits U.S.-based issuing authorities grounded in ISO/IEC 17024:2012.¹ Compliance to this standard has become essential within the cybersecurity domain. This is due, in part, to DoD 8570.01-M,⁴ which di-

rects employees of certain functions to obtain cyber-security certification. These certifications must be from ISO/IEC 17024-compliant issuers. The effect of this directive can be seen in *Certification Demand* by the presence of the entry "IAT Level II Certification."

Seidman and Kruchten both advocate for stakeholders to seek compliance with this standard as well.^{10,11} ISO/IEC 17024:2012 provides development and maintenance requirements for personal certification programs. The standard seeks to address assurance that a certificate holder meets the requirements of the certification, and that assessment and periodic reassessment of competence follows globally accepted processes.⁷

Underpinning each certification is a unique competency model. This model states the cognitive or knowledge-processing level expected for each component of the body of knowledge and level of skill proficiency.^{11,16} The model describes the means by which a candidate demonstrates required competencies and how performance is evaluated.¹¹ This model instills coherency within the design, development, testing, and delivery of assessment. The educational aspects of certifications should align with the model as well. Although developed independently, this common alignment improves the likelihood that educational content and services adequately prepare a qualified candidate for the upcoming assessment. Any stipulations for pre-certification qualifications are formally presented in the competency model.¹¹ These pre-certification qualifications function as a benchmark by which to assess prior achievements, and they help to characterize the expected audience when developing assessment and training materials.

Generally, the content scope of a certification often consists of a topic-oriented body of knowledge and possibly a code of professional behavior. Technology-vendor certifications tend to focus on their technology as the sole orientation of their certifications. Professional-practice certifications are incomplete without a code of professional behavior, which often consists of a code of ethics and accepted standards of practices.^{11,15,17}

To extrapolate from Seidman's dis-

cussion on software engineering certification, the code of ethics outlines the moral principles and ideals for socially responsible performance of professional responsibilities relevant to a line of work.¹¹ The body of knowledge is a compiled set of understandings, such as facts, concepts, principles, tools, measures, methods, and goals. Useful bodies of knowledge, such as Software Engineering Body of Knowledge ISO/IEC TR 19759:2005, are drawn from established understandings that are found to be relevant and reliable by stakeholders invested in a topic. A certification's relevance and perceived practical competency is strongly influenced by the process used and contributors involved to form the content scope.

Costs

Achievement costs are related to the preparation and completion of certification assessments. Certifications require preparation time. Certifications require preparation resources, such as practice exams, books, and training courses. Many certification exams can be taken at local commercial testing centers, but assessment availability may require the candidate to incur travel expenses. Finally, there are the exam fees. As part of this cost item, there may be an application fee in addition to the exam fee. Exam fees are established to cover testing-center expenses as well as exam authoring, quality, and maintenance costs. Assessment grading may be automated; however, greater costs should be expected if grading requires human involvement. Practice or skills-based assessments requiring specialized equipment have fees to cover the capital and operational expenses necessary to supply the assessment environment.

Many certifications are designed with recertification requirements to be completed within defined renewal cycles. Recertification often involves maintenance costs, such as annual maintenance fees with the issuer and/or possibly a recertification fee charged upon renewal. Vendor- or technology-centric certifications effectively expire upon new major product releases, requiring the passing of additional assessments tailored to the new release in order to

maintain currency. Recertification for vendor- or technology-neutral certifications typically involves ongoing education. This education requirement may incur education fees depending on availability, types, and quality of education stipulated by the issuing authority. A holder may be able to apply direct professional practice to recertification. Conferences and formal instruction offerings are likely to impose a fee. Non-local in-person offerings incur travel expenses.

Although an education opportunity is provided at no cost, the time to attend is unavoidable. If one holds multiple certifications, the chances of non-overlapping recertification requirements is high. To diversify the types of education sought within a recertification cycle, an issuer may specify the number of times a certain education type or topic area may be applied. Recertification costs may overtake achievement costs in the long run.

Opportunity costs are intangible and yet real. The most apparent cost item is work-life balance. The time spent preparing for and maintaining a certification is often in addition to work and personal demands. Within any popular category of certifications are multiple options from which a practitioner must choose. Given the financial outlay and work-life balance issues, for some it may be a burden to hold numerous certifications within closely related or overlapping focus areas.

Professional mobility is another opportunity cost to consider. Although Seidman raises mobility as an issue with regards to software engineering,¹¹ it is reasonable to broadly consider that employers within a particular country may not recognize a particular localized issuer, competency model, professional practices, the body of knowledge, or the assessment as being relevant in their country for their purposes. While certifications may be cheaper than a college degree in the short term, not pursuing a post-secondary or graduate degree may have its own opportunity costs regarding advancement or career flexibility.

Closing Thoughts

In many countries, professional certifications provide credible evidence of

a practitioner's competence⁴ in areas of focus. In the U.S., the value placed on certifications is not consistent across computer occupations. Consider the certification-demand coherence of an occupation subcategory when investing in a certification. Low coherence is a sign that receptiveness to the achievement of a certification is more strongly linked to an employer's explicit expectations. Recognition for an unrequested certification may be too low to make a favorable impression. If the opportunity presents itself, advocating for the worthiness of a relevant but unrecognized certification might help tip the value calculus to a favorable value determination. When seeking to diversify one's career, recent role-relevant experience may still be necessary beyond achieving a relevant certification.

When choosing a new certification, consider one's desired occupation. The value of achieving a particular certification is strongly influenced by what employers think about that credential. Greater demand coherence appears to indicate a lower risk for choosing a certification prior to focusing on a particular job opportunity. This should hold true on the condition that the choice falls within the current set of popular certifications. Consider investing in ISO 17024-compliant certifications; this standard is gaining in stature as a quality criterion. An ISO 17024-compliant certification counters the perception of non-compliance as being a deficiency. Recertification should be embraced. The education required for recertification encourages currency and familiarity with the focus area.

Investment in many quality certifications continues after achieving the certification. Recertification requires time and money commitments within each recertification period. A good time to reevaluate a certification investment is at the time of renewal. Some questions to consider are: What is one's level of satisfaction with one's present role and related career path? What alternative certifications with higher assessor values could replace the present certification?

However, do not let a certification expire if it is vital to one's livelihood; this may result in job loss—for exam-

ple, DoD 8570.01-M policy⁴—and reputational damage. Restoring an expired certification may require satisfying the current assessment requirements.

Taking on constructive challenges and new learning are both valuable for personal growth, insight, and confidence. Although personally valuable, certifications are subject to market perception, assessor value judgments, quality considerations, and personal costs. One should be clear about each certification's contribution to one's career prospects and related post-achievement commitment. ■

References

1. ANSI Personnel Certification. <https://anab.ansi.org/credentialing/personnel-certification>.
2. Computer and information technology occupations. *U.S. Bureau of Labor Statistics*. <https://www.bls.gov/ooh/computer-and-information-technology/print/home.htm>.
3. Cunningham, Evan. Professional certifications and occupational licenses: Evidence from the Current Population Survey. *Monthly Labor Review* (2019).
4. Information assurance workforce improvement program. *Department of Defense* (Nov. 2015). <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodm/857001m.pdf>.
5. Dreyfus, S. and Dreyfus, H. A five-stage model of the mental activities involved in direct skill acquisition. *Univ. Cal. Berkley Operations Research Center* (Feb. 1980).
6. Information security and financial institutions: An FTC workshop to examine GLB safeguards. *Federal Trade Commission* (July 2020). https://www.ftc.gov/system/files/documents/public_events/1567141/slides-glb-workshop.pdf.
7. ISO/IEC 17024:2012(en) Conformity assessment—General requirements for bodies operating certification of persons. *ISO*. <https://www.iso.org/obp/ui/#iso:std:iso-iec:17024:ed-2:v1:en>.
8. Karlin, S. Certification uncertainty. *IEEE Spectrum* (Nov. 2006).
9. Knowles, W., Such, J., Gougliadis, A., Misra, G., and Rashid, A. All that glitters is not gold: On the effectiveness of cybersecurity qualifications. *IEEE Computer* (Dec. 2017), 60-71.
10. Kruchten, P. Lifelong learning for lifelong employment. *IEEE Software* (Jul./Aug. 2015), 85-87.
11. Seidman, S. Software engineering certification schemes. *IEEE Computer* (May 2008), 87-89.
12. Seidman, S. An international perspective on professional software engineering credentials. *Software Engineering: Effective Teaching and Learning Approaches and Practices*. IGI Global, eds. H.J.C. Ellis, S.A. Demurjian, and J.F. Naveda (2009), Ch. 18.
13. Seidman, S. Computing: An emerging profession? *ACM Inroads* 5 (2014), 6-11.
14. IT training. <https://trainingindustry.com/wiki/it-and-technical-training/it-training/>.
15. Trippi, L. Benefits of certification. *IEEE Computer* (June 2002), 31-33.
16. Walrad, C., Lane, M., Walk, J., and Hirst, D. Architecting a profession. *IEEE IT Pro* (Jan./Feb. 2014), 42-49.
17. Walrad, C. Standards for the Enterprise IT Profession. *IEEE Computer* (Mar. 2017), 70-73.

Mark Tannian holds CISSP and PMP certifications. He is a faculty member at St. John's University in Queens, NY, USA, and also offers professional development services.

Willie Coston IV is an Information Technology sales manager at TEKsystems Inc. in New York, NY, USA. His expertise is in technology services and leading large-scale digital transformation efforts.

Copyright held by author/owner.
Publication rights licensed to ACM.

The pursuit of responsible AI raises the ante on both the trustworthy computing and formal methods communities.

BY JEANNETTE M. WING

Trustworthy AI

FOR CERTAIN TASKS, AI systems have achieved good enough performance to be deployed in our streets and our homes. Object recognition helps modern cars see. Speech recognition helps personalized voice assistants, such as Siri and Alexa, converse. For other tasks, AI systems have even exceeded human performance. AlphaGo was the first computer program to beat the best Go player in the world.

The promise of AI is huge. They will drive our cars. They will help doctors diagnose disease more accurately.⁵⁴ They will help judges make more consistent court decisions. They will help employers hire more suitable job candidates.

However, we know these AI systems can be brittle and unfair. Adding graffiti to a stop sign fools the classifier into saying it is not a stop sign.²² Adding noise to an image of a benign skin lesion fools the classifier into saying it is malignant.²³ Risk assessment tools used in U.S. courts have shown to be biased against blacks.⁴ Corporate recruiting tools have been shown to be biased against women.¹⁷

How then can we deliver on the promise of the benefits of AI but address these scenarios that have

life-critical consequences for people and society? In short, how can we achieve *trustworthy AI*?

The ultimate purpose of this article is to rally the computing community to support a broad-based, long-term research program on trustworthy AI, drawing on the expertise and sensibilities from multiple research communities and stakeholders. This article focuses on addressing three key research communities because: trustworthy AI adds new desired properties above and beyond those for *trustworthy computing*; AI systems require new *formal methods* techniques, and in particular, the role of data raises brand new research questions; and AI systems can likely benefit from the scrutiny of formal methods for ensuring trustworthiness. By bringing together researchers in trustworthy computing, formal methods, and AI, we aim to foster a new research community across academia, industry, and government in trustworthy AI.

From Trustworthy Computing to Trustworthy AI

The landmark *Trust in Cyberspace* 1999 National Academies report lay the foundations of trustworthy computing and what continues to be an active research area.⁴¹

Around the same time, the National Science Foundation started a series of programs on trust. Starting with

» key insights

- The set of trustworthiness properties for AI systems, in contrast to traditional computing systems, needs to be extended beyond reliability, security, privacy, and usability to include properties such as probabilistic accuracy under uncertainty, fairness, robustness, accountability, and explainability.
- To help ensure their trustworthiness, AI systems can benefit from the scrutiny of formal methods.
- AI systems raise the bar on formal methods for two key reasons: the inherent probabilistic nature of machine-learned models, and the critical role of data in training, testing, and deploying a machine-learned model.



IMAGE BY ANDRZJ BORYS ASSOCIATES, USING SHUTTERSTOCK

Trusted Computing (initiated in 2001), then Cyber Trust (2004), then Trustworthy Computing (2007), and now Secure and Trustworthy Cyberspace (2011), the Computer and Information Science and Engineering Directorate has grown the academic research community in trustworthy computing. Although it started within the computer science community, support for research in trustworthy computing now spans multiple directorates at NSF and engages many other funding organizations, including, through the Networking and Information Technology Research and Development (NITRD) Program, 20 federal agencies.

Industry has also been a leader and active participant in trustworthy computing. With Bill Gates's January 2002

"Trustworthy Computing" memo,²⁶ Microsoft signaled to its employees, customers, shareholders, and the rest of the information technology sector the importance of trustworthy software and hardware products. It referred to an internal Microsoft white paper, which identified four pillars to trustworthiness: security, privacy, reliability, and business integrity.

After two decades of investment and advances in research and development, *trustworthy* has come to mean a set of (overlapping) properties:

- ▶ Reliability: Does the system do the right thing?
- ▶ Safety: Does the system do no harm?
- ▶ Security: How vulnerable is the system to attack?

- ▶ Privacy: Does the system protect a person's identity and data?

- ▶ Availability: Is the system up when I need to access it?

- ▶ Usability: Can a human use it easily?

The *computing* systems for which we want such properties to hold are hardware and software systems, including their interaction with humans and the physical world. Academia and industry have made huge strides in trustworthy computing in the past decades. However, as technology advances and as adversaries get more sophisticated, trustworthy computing remains a holy grail.

AI systems raise the bar in terms of the set of properties of interest. In addition to the properties associated with trustworthy computing (as noted),

we also want (overlapping) properties such as:

- **Accuracy:** How well does the AI system do on new (unseen) data compared to data on which it was trained and tested?

- **Robustness:** How sensitive is the system's outcome to a change in the input?

- **Fairness:** Are the system outcomes unbiased?

- **Accountability:** Who or what is responsible for the system's outcome?

- **Transparency:** Is it clear to an external observer how the system's outcome was produced?

- **Interpretability/Explainability:** Can the system's outcome be justified with an explanation that a human can understand and/or that is meaningful to the end user?

- **Ethical:** Was the data collected in an ethical manner? Will the system's outcome be used in an ethical manner?

- ...and others, yet to be identified

The machine learning community considers accuracy as a gold standard, but trustworthy AI requires us to explore trade-offs among these properties. For example, perhaps we are willing to give up on some accuracy in order to deploy a fairer model. Also, some of the above properties may have different interpretations, leading to different formalizations. For example, there are many reasonable notions of fairness,⁴⁰ including demographic parity, equal odds, and individual fairness,²⁰ some of which are incompatible with each other.^{12,33}

Traditional software and hardware systems are complex due to their size and the number of interactions among their components. For the most part, we can define their behavior in terms of discrete logic and as deterministic state machines.

Today's AI systems, especially those using deep neural networks, add a dimension of complexity to traditional computing systems. This complexity is due to their inherent probabilistic nature. Through probabilities, AI systems model the uncertainty of human behavior and the uncertainty of the physical world. More recent advances in machine learning, which rely on big data, add to their probabilistic nature, as data from the real world are just points in a probability space. Thus,

trustworthy AI necessarily directs our attention from the primarily deterministic nature of traditional computing systems to the probabilistic nature of AI systems.

Verify, to Trust

How can we design, implement, and deploy AI systems to be trustworthy?

One approach for building end-user trust in computing systems is formal verification, where properties are proven once and for all over a large domain, for example, for all inputs to a program or for all behaviors of a concurrent or distributed system. Alternatively, the verification process identifies a counterexample, for example, an input value where the program produces the wrong output or a behavior that fails to satisfy the desired property, and thus provides valuable feedback on how to improve the system. Formal verification has the advantage of obviating the need to test individual input values or behaviors one-by-one, which for large (or infinite) state spaces is impossible to achieve completely. Early success stories in formal methods, for example, in verifying cache coherence protocols⁴⁸ and in detecting device driver bugs,⁵ led to their scalability and practicality today. These approaches are now used in the hardware and software industry, for example, Intel,²⁹ IBM,⁶ Microsoft,⁵ and Amazon.^{15,44} Due to advances in formal methods languages, algorithms, and tools, and to the increased scale and complexity of hardware and software, we have seen in the past few years a new surge of interest and excitement in formal verification, especially for ensuring the correctness of critical components of system infrastructure.^{7,10,11,15,27,30,34,49}

Formal verification is a way to provide provable guarantees and thus increase one's trust that the system will behave as desired.

From traditional formal methods to formal methods for AI. In traditional formal methods, we want to show that a model M *satisfies* (\models) a property P .

$$M \models P$$

M is the object to be verified—be it a program or an abstract model of a complex system, for example, a concurrent, distributed, or reactive system. P is the correctness property, expressed in some discrete logic. For example, M

might be a concurrent program that uses locks for synchronization and P might be “deadlock free.” A proof that M is deadlock free means any user of M is assured that M will never reach a deadlocked state. To prove that M satisfies P , we use formal mathematical logics, which are the basis of today's scalable and practical verification tools such as model checkers, theorem provers, and satisfiability modulo theories (SMT) solvers.

Especially when M is a concurrent, distributed, or reactive system, in traditional formal methods, we often add explicitly a specification of a system's environment E in the formulation of the verification task:

$$E, M \models P$$

For example, if M is a parallel process, E might be another process with which M interacts (and then we might write $E \parallel M \models P$, where \parallel stands for parallel composition). Or, if M is device driver code, E might be a model of the operating system. Or, if M is a control system, E might be a model of its environment that closes the control loop. The specification of E is written to make explicit the assumptions about the environment in which the system is to be verified.

For verifying AI systems, M could be interpreted to be a complex system, for example, a self-driving car, within which is a component that is a machine-learned model, for example, a computer vision system. Here, we would want to prove P , for example, safety or robustness, with respect to M (the car) in the context of E (traffic, roads, pedestrians, buildings, and so on). We can view proving P as proving a “system-level” property. Seshia et al. elaborate on the formal specification challenges with this perspective,⁵¹ where a deep neural network might be a black-box component of the system M .

But what can we assert about the machine learned model, for example, a DNN, that is a critical component of the system? Is there a robustness or fairness property we can verify of the machine-learned model itself? Are there white-box verification techniques that can take advantage of the structure of the machine learned model? Answering these questions raises new verification challenges.


Verifying a machine-learned model M.

For verifying an ML model, we reinterpret M and P: M stands for a machine-learned model. P stands for a trustworthy property, for example, safety, robustness, privacy, or fairness.


Verifying AI systems ups the ante over traditional formal methods. There are two key differences: the inherent probabilistic nature of the machine-learned model and the role of data.

The inherent probabilistic nature of M and P, and thus the need for probabilistic reasoning (=). The ML model, M, itself is semantically and structurally different from a typical computer program. As mentioned, it is inherently probabilistic, taking inputs from the real world, that are perhaps mathematically modeled as a stochastic process, and producing outputs that are associated with probabilities. Internally, the model itself operates over probabilities; for example, labels on edges in a deep neural network are probabilities and nodes compute functions over these probabilities. Structurally, because a machine generated the ML model, M itself is not necessarily something human readable or comprehensible; crudely, a DNN is a complex structure of if-then-else statements that would unlikely ever be written by a human. This “intermediate code” representation opens up new lines of research in program analysis.

The properties P themselves may be formulated over continuous, not (just) discrete domains, and/or using expressions from probability and statistics. Robustness properties for deep neural networks are characterized as predicates over continuous variables.¹⁸ Fairness properties are characterized in terms of expectations with respect to a loss function over reals (for example, see Dwork et al.²⁰). Differential privacy is defined in terms of a difference in probabilities with respect to a (small) real value.²¹ Note that just as with properties such as usability for trustworthy computing, some desired properties of trustworthy AI systems, for example, transparency or ethics, have yet to be formalized or may not be formalizable. For such properties, a framework that considers legal, policy, behavioral and social rules and norms could provide the context within which a formalizable question can be answered. In



Formal verification is a way to provide provable guarantees and thus increase one's trust that the system will behave as desired.



short, verification of AI systems will be limited to what can be formalized.

These inherently probabilistic models M and associated desired trust properties P call for scalable and/or new verification techniques that work over reals, non-linear functions, probability distributions, stochastic processes, and so on. One stepping-stone to verifying AI systems is probabilistic logics and hybrid logics (for example, Alur et al.,³ Kwiatkowska et al.³⁵ and Platzer⁴⁶), used by the cyber-physical systems community. Another approach is to integrate temporal logic specifications directly in reinforcement learning algorithms.²⁴ Even more challenging is that these verification techniques need to operate over machine-generated code, in particular code that itself might not be produced deterministically.^a

The role of data. Perhaps the more significant key difference between traditional formal verification and verification for AI systems is the role of data—data used in training, testing, and deploying ML models. Today's ML models are built and used with respect to a set, D, of data. For verifying an ML model, we propose to make explicit the assumptions about this data, and formulate the verification problem as:

$$D, M \models P$$

Data is divided into *available data* and *unseen data*, where *available data* is data-at-hand, used for training and testing M; and *unseen data* is data over which M needs (or is expected) to operate without having seen it before. The whole idea behind building M is so that based on the data on which it was trained and tested, M would be able to make predictions on data it has never seen before, typically to some degree of accuracy.

Making the role of data explicit raises novel specification and verification challenges, roughly broken into these categories, with related research questions:

Collection and partitioning of available data:

- How much data suffices to build

^a The ways in which machine learning models, some with millions of parameters, are constructed today, perhaps through weeks of training on clusters of CPUs, TPUs, and GPUs, raise a meta-issue of trust: scientific reproducibility.


a model M for a given property P ? The success of deep learning has taught us that with respect to accuracy, the more data, the better the model, but what about other properties? Does adding more data to train or test M make it more robust, fairer, and so on, or does it not have an effect with respect to the property P ? What new kind of data needs to be collected if a desired property does not hold?

► How do we partition an available (given) dataset into a training set and a test set? What guarantees can we make of this partition with respect to a desired property P , in building a model M ? Would we split the data differently if we were training the model with respect to multiple properties at the same time? Would we split the data differently if we were willing to trade one property over another?

Specifying unseen data: Including D in the formal methods framework $D, M \models P$ gives us the opportunity to state explicitly assumptions about the unseen data.

► How do we specify the data and/or characterize properties of the data? For example, we could specify D as a stochastic process that generates inputs over which the ML model needs to be verified. Or, we could specify D as a data distribution. For a common statistical model, for example, a normal distribution, we could specify D in terms of its parameters, for example, mean and variance. Probabilistic programming languages, for example, Stan,⁸ might be a starting point for specifying statistical models. But what of large real-world datasets that do not fit common statistical models, or which have thousands of parameters?

► In specifying unseen data, by definition, we will need to make certain assumptions about the unseen data. Would these assumptions not then be the same as those we would make to build the model M in the first place? More to the point: *How can we trust the specification of D ?* This seemingly logical deadlock is analogous to the problem in traditional verification, where given an M , we need to assume the specifications of the elements E and P are “correct” in the verification task $E, M \models P$. Then in the verification process, we may need to modify E and/or P (or even M). To break the circular



The formal methods community has recently been exploring robustness properties of AI systems, in particular, image processing systems used in autonomous vehicles.



reasoning at hand, one approach is to use a different validation approach for checking the specification of D ; such approaches could borrow from a repertoire of statistical tools. Another approach would be to assume an initial specification is small or simple enough that it can be checked by (say, manual) inspection; then we use this specification to bootstrap an iterative refinement process. (We draw inspiration from the counterexample guided abstraction and refinement method¹⁴ of formal methods.) This refinement process may necessitate modifying D , M , and/or P .

► How does the specification of unseen data relate to the specification of the data on which M was trained and tested?

In traditional verification, we aim to prove property, P , a universally quantified statement: for example, *for all* input values of integer variable x , the program will return a positive integer; or *for all* execution sequences x , the system will not deadlock.

So, the first question for proving P of an ML model, M , is: in P , what do we quantify over? For an ML model that is to be deployed in the real world, one reasonable answer is to quantify over data distributions. But a ML model is meant to work only for certain distributions that are formed by real world phenomena, and *not* for arbitrary distributions. We do not want to prove a property *for all* data distributions. This insight on the difference in what we quantify over and what the data represents for proving a trust property for M leads to this novel specification question:

► How can we specify the class of distributions over which P should hold for a given M ? Consider robustness and fairness as two examples:

► For robustness, in the adversarial machine learning setting, we might want to show that M is robust to all norm-bounded perturbations D . More interestingly, we might want to show M is robust to all “semantic” or “structural” perturbations for the task at hand. For example, for some vision tasks, we want to consider rotating or darkening an image, not just changing any old pixel.

► For fairness, we might want to show the ML model is fair on a given dataset and all unseen datasets that are

“similar” (for some formal notion of “similar”). Training a recruiting tool to decide whom to interview on one population of applicants should ideally be fair on any future population. How can we specify these related distributions?

Toward building a fair classifier that is also robust, Mandal et al. show how to adapt an online learning algorithm that finds a classifier that is fair over a class of input distributions.³⁷

Verification task: Once we have a specification of D and P , given an M , we are then left with verifying that M satisfies P , given any assumptions we have made explicit about available and unseen data in D , using whatever logical framework (\models) we have at hand.

► How do we check the available data for desired properties? For example, if we want to detect whether a dataset is fair or not, what should we be checking about the dataset?

► If we detect the property does not hold, how do we fix the model, amend the property, or decide what new data to collect for retraining the model? In traditional verification, producing a counterexample, for example, an execution path that does not satisfy P , helps engineers debug their systems and/or designs. What is the equivalent of a “counterexample” in the verification of an ML model and how do we use it?

► How do we exploit the explicit specification of unseen data to aid in the verification task? Just as making explicit the specification of the environment, E , in the verification task $E, M \models P$, how can we leverage having an explicit specification of D ?

► How can we extend standard verification techniques to operate over data distributions, perhaps taking advantage of the ways in which we formally specify unseen data?

These two key differences—the inherent probabilistic nature of M and the role of data D —provide research opportunities for the formal methods community to advance specification and verification techniques for AI systems.

Related work. The formal methods community has recently been exploring robustness properties of AI systems,¹⁸ in particular, image processing systems used in autonomous vehicles. The state-of-the-art VerifAI system¹⁹ explores the verification of robustness

of autonomous vehicles, relying on simulation to identify execution traces where a cyber-physical system (for example, a self-driving car) whose control relies on an embedded ML model could go awry. Tools such as ReluVal⁵⁶ and Neurify⁵⁷ look at robustness of DNNs, especially as applied to safety of autonomous vehicles, including self-driving cars and aircraft collision avoidance systems. These tools rely on interval analysis as a way to cut down on state exploration, while still providing strong guarantees. A case study using Verisig to verify the safety of a DNN-based controller for the F1/10 racing car platform provides a benchmark for comparing different DNN configurations and sizes of input data and identifies a current gap between simulation and verification.³²

FairSquare² uses probabilistic verification to verify fairness of ML models. LightDP⁶⁰ transforms a probabilistic program into a non-probabilistic one, and then does type inference to automate verification of privacy budgets for differential privacy.

These pieces of work are in the spirit of trustworthy AI, but each focuses on only one trust property. Scaling their underlying verification techniques to industry-scale systems is still a challenge.

Additional formal methods opportunities. Today’s AI systems are developed to perform a particular task in mind, for example, face recognition or playing Go. How do we take into consideration the task that the deployed ML model is to perform in the specification and verification problem? For example, consider showing the robustness of a ML model, M , that does image analysis: For the task of identifying cars on the road, we would want M to be robust to the image of any car that has a dent in its side; but for the task of quality control in an automobile manufacturing line, we would not.

Previously, we focused on the verification task in formal methods. But the machinery of formal methods has also successfully been used recently for program synthesis.²⁸ Rather than post-facto verification of a model M , can we develop a “correct-by-construction” approach in building M in the first place? For example, could we add the desired trustworthy property, P , as a constraint as we train and test M , with the inten-

tion of guaranteeing that P holds (perhaps for a given dataset or for a class of distributions) at deployment time? A variant of this approach is to guide the ML algorithm design process by checking at each step that the algorithm never satisfies an undesirable behavior.⁵³ Similarly, safe reinforcement learning addresses learning policies in decision processes where safety is added as a factor in optimization or an external constraint in exploration.²⁵

The laundry list of properties enumerated at the outset of this article for trustworthy AI is unwieldy, but each is critical toward building trust. A task ahead for the research community is to formulate commonalities across these properties, which can then be specified in a common logical framework, akin to using temporal logic^{38,47} for specifying safety (“nothing bad happens”) and liveness (“something good eventually happens”) properties³⁶ for reasoning about correctness properties of concurrent and distributed systems.

Compositional reasoning enables us to do verification on large and complex systems. How does verifying a component of an AI system for a property “lift” to showing that property holds for the system? Conversely, how does one decompose an AI system into pieces, verify each with respect to a given property, and assert the property holds of the whole? Which properties are global (elude compositionality) and which are local? Decades of research in formal methods for compositional specification and verification give us a vocabulary and framework as a good starting point.

Statistics has a rich history in model checking^b and model evaluation, using tools such as sensitivity analysis, prediction scoring, predictive checking, residual analysis, and model criticism. With the goal of validating an ML model satisfies a desired property, these statistical approaches can complement formal verification approaches, just as testing and simulation complement verification of computational systems. Even more relevantly, as mentioned in “The role of data” noted earlier, they can help with the evaluation of any sta-

^b Not to be confused with formal method’s notion of model checking, where a finite state machine (computational model of a system) is checked against a given property specification.^{13,50}

tistical model used to specify unseen data, D , in the $D, M \neq P$ problem. An opportunity for the formal methods community is to combine these statistical techniques with traditional verification techniques (for early work on such a combination, see Younes et al.⁵⁹).

Building a Trustworthy AI Community

Just as for trustworthy computing, formal methods is only one approach toward ensuring increased trust in AI systems. The community needs to explore many approaches, especially in combination, to achieve trustworthy AI. Other approaches include testing, simulation, run-time monitoring, threat modeling, vulnerability analysis, and the equivalent of design and code reviews for code and data. Moreover, besides technical challenges, there are societal, policy, legal, and ethical challenges.

On October 30–November 1, 2019, Columbia University’s Data Science Institute hosted an inaugural Symposium on Trustworthy AI¹ sponsored by Capital One, a DSI industry affiliate. It brought together researchers from formal methods, security and privacy, fairness, and machine learning. Speakers from industry brought a reality check to the kinds of questions and approaches the academic community are pursuing. The participants identified research challenge areas, including:

- ▶ Specification and verification techniques;
- ▶ “Correctness-by-construction” techniques;
- ▶ New threat models and system-level adversarial attacks;
- ▶ Processes for auditing AI systems that consider properties such as explainability, transparency, and responsibility;
- ▶ Ways to detect bias and de-bias data, machine learning algorithms, and their outputs;
- ▶ Systems infrastructure for experimenting for trustworthiness properties;
- ▶ Understanding the human element, for example, where the machine is influencing human behavior; and
- ▶ Understanding the societal element, including social welfare, social norms, morality, ethics, and law.

Technology companies, many of which push the frontiers of machine learning and AI, have not been sitting still. They realize the importance

of trustworthy AI for their customers, their business, and social good. Of predominant concern is fairness. IBM’s AI Fairness 360 provides an open source toolkit to check for unwanted bias in datasets and machine learning models.⁵⁵ Google’s TensorFlow kit provides “fairness indicators” for evaluating binary and multi-class classifiers for fairness.³¹ Microsoft’s Fairlearn is an open source package for machine learning developers to assess their systems’ fairness and to mitigate observed unfairness.³⁹ At its F8 conference in 2018, Facebook announced its Fairness Flow tool intended “to measure for potential biases for or against particular groups of people.”⁵² In the spirit of industry and government collaborations, Amazon and the National Science Foundation have partnered since 2019 to fund a “Fairness in AI” program.⁴³

In 2016, DARPA focused on explainability by launching the Explainable AI (XAI) Program.¹⁶ The goal of this program was to develop new machine learning systems that could “explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future.” With explainability would come increased trust by an end user to believe and adopt the outcome of the system.

Through the Secure and Trustworthy Cyberspace Program, NSF funds a Center on Trustworthy Machine Learning⁹ led by Penn State University and involving researchers from Stanford, UC Berkeley, UC San Diego, University of Virginia, and University of Wisconsin. Their primary focus is on addressing adversarial machine learning, complementary to the formal methods approach outlined previously. (In the interests of full disclosure, the author is on this Center’s Advisory Board.) In October 2019, the National Science Foundation announced a new program to fund National AI Institutes.⁴² One of the six themes it named was “Trustworthy AI,” emphasizing properties such as reliability, explainability, privacy, and fairness.

The NITRD report on AI and cybersecurity calls explicitly for research in the specification and verification of AI systems and for trustworthy AI decision-making.⁴⁵ Finally, in December 2020, the White House signed an ex-

ecutive order on trustworthy AI to provide guidance to U.S. federal agencies in adopting AI for their services and to foster public trust in AI.⁵⁸

Just as for trustworthy computing, government, academia, and industry are coming together to drive a new research agenda in trustworthy AI. We are upping the ante on a holy grail!

Acknowledgments

During 2002–2003, I was fortunate to spend a sabbatical at Microsoft Research and witnessed firsthand how trustworthy computing permeated the company. It was also the year when the SLAM project⁵ showed how the use of formal methods could systematically detect bugs in device driver code, which at the time was responsible for a significant fraction of “blue screens of death.” Whereas formal methods had already been shown to be useful and scalable for the hardware industry, the SLAM work was the first industry-scale project that showed the effectiveness of formal methods for software systems. I also had the privilege to serve on the Microsoft Trustworthy Computing Academic Advisory Board from 2003–2007 and 2010–2012.

When I joined NSF in 2007 as the Assistant Director for the Computer and Information Science and Engineering Directorate, I promoted trustworthy computing across the directorate and with other federal agencies via NITRD. I would like to acknowledge my predecessor and successor CISE ADs, and all the NSF and NITRD program managers who cultivated the community in trustworthy computing. It is especially gratifying to see how the Trustworthy Computing program has grown to the Secure and Trustworthy Cyberspace program, which continues to this day.

ACM sponsors the annual FAT* conference, which originally promoted fairness, accountability, and transparency in machine learning. The Microsoft Research FATE group added “E” for ethics. FAT* has since grown to recognize other properties, including ethics, as well as the desirability of these properties for AI systems more generally, not just machine learning. My list of trustworthy AI properties is inspired by this community.

I would like to acknowledge S. Agrawal, R. Geambasu, D. Hsu, and S. Jana for

their insights into what makes verifying AI systems different from verifying traditional computing systems. Thanks to A. Chaintreau for instigating my journey on trustworthy AI. Special thanks to R. Geambasu and T. Zheng who provided comments on an earlier draft of this article. Thanks also to the anonymous reviewers for their pointers to relevant related work.

Final thanks to Capital One, JP Morgan, the National Science Foundation, and the Sloan Foundation for their support and encouragement to promote trustworthy AI. □

References

1. Agrawal, S. and Wing, J.M. Trustworthy AI Symposium. Columbia University, (Oct. 30–Nov. 1, 2019); <https://datascience.columbia.edu/trustworthy-ai-symposium>.
2. Albarghouthi, A., D'Antoni, L., Drews, S. and Nori, A. FairSquare: Probabilistic verification of program fairness. In *Proceedings of ACM OOPSLA '17*.
3. Alur, R., Henzinger, T.A. and Ho, P.H. Automatic symbolic verification of embedded systems. *IEEE Trans. Software Eng.* 22 (1996), 181–201.
4. Angwin, J., Larson, J., Mattu, S. and Kirchner, L. Machine bias. *ProPublica* (May 23, 2016).
5. Ball, T., Cook, B., Levin, V. and Rajamani, S. SLAM and Static driver verifier: Technology Transfer of formal methods inside Microsoft. Technical Report MSR-TR-2004-08. Microsoft Research, Jan. 2004.
6. Baumgartner, J. Integrating formal verification into mainstream verification: The IBM experience. *Formal Methods in Computer-Aided Design*, Haifa, Israel, 2006.
7. Bhargavan, K. et al. Everest: Towards a verified, drop-in replacement of HTTPS. In *Proceedings of the 2nd Summit on Advances in Programming Languages*, May 2017.
8. Carpenter, B. et al. Stan: A probabilistic programming language. *J. Statistical Software* 76, 1 (2017); DOI 10.18637/jss.v076.i01
9. Center for Trustworthy Machine Learning; <https://ctml.psu.edu>
10. Chen, H., Chajed, T., Konradi, A., Wang, S., Ileri, A., Chlipala, A., Kaashoek, M.F. and N. Zeldovich, N. Verifying a high-performance crash-safe file system using a tree specification. In *Proceedings of the 26th ACM Symposium on Operating Systems Principles*, 2017.
11. Chen, H., Ziegler, D., Chajed, T., Chlipala, A., Kaashoek, M.F. and Zeldovich, N. Using crash Hoare logic for certifying the FSCQ file system. In *Proceedings of the 25th ACM Symp. Operating Systems Principles*, 2015.
12. Chouldechova, A. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. In *Proceedings of FATML*, 2016.
13. Clarke, E.M. and Emerson, E.A. Characterizing correctness properties of parallel programs using fixpoints. *Automata, Languages and Programming, Lecture Notes in Computer Science* 85 (1980), 169–181
14. Clarke, E., Grumberg, O., Jha, S., Lu, Y. and Veith, H. Counterexample-guided abstraction refinement. *Computer Aided Verification*. E.A. Emerson, A.P. Sistla, eds. *Lecture Notes in Computer Science* 1855 (2000). Springer, Berlin, Heidelberg.
15. Cook, B. Formal reasoning about the security of Amazon Web Services. *Proceedings of the International Conference on Computer Aided Verification*, Volume 10981, 2018.
16. DARPA. Explainable AI (XAI) Program. Matt Turek, Defense Advanced Research Projects Agency, 2016; <https://www.darpa.mil/program/explainable-artificial-intelligence>.
17. Dastin, J. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, Oct. 9, 2018.
18. Dreossi, T., Ghosh, S., Sangiovanni-Vincentelli, A.L. and Seshia, S.A. A formalization of robustness for deep neural networks. In *Proceedings of the AAAI Spring Symp. Workshop on Verification of Neural Networks*, Mar. 2019.
19. Dreossi, T., Ghosh, S., Sangiovanni-Vincentelli, A.L. and Seshia, S.A. VERIFAI: A toolkit for the formal design and analysis of artificial intelligence-based systems. In *Proceedings of Intern. Conf. Computer-Aided Design*, 2019.
20. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. and Zemel, R. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 2012; <https://doi.org/10.1145/2090236.2090255>
21. Dwork, C., McSherry, F., Nissim, K. and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Con. Theory of Cryptography*. S. Halevi and T. Rabin, Eds. Springer-Verlag, Berlin, Heidelberg, 2006, 265–284; DOI:10.1007/11681878_14
22. Eykholt, K. et al. Robust physical-world attacks on deep learning visual classification. In *Proceedings of CVPR 2017*.
23. Finlayson, S.G., Bowers, J.D., Ito, J., Zittrain, J.L., Beam, A.L., Kohane, I.S. Adversarial attacks on medical machine learning. *Science* 363, 6433 (2019), 1287–1289; DOI: 10.1126/science.aaw4399
24. Gao, Q., Hajinezhad, D., Zhang, Y., Kantaros, Y. and Zavlanos, M.M. Reduced variance deep reinforcement learning with temporal logic specifications. In *Proceedings of ACM/IEEE Intern. Conf. Cyber-Physical Systems*, 2019, 237–248.
25. Garcia, J. and Fernandez, F. A comprehensive survey on safe reinforcement learning. *J. Machine Learning Research* 16 (2015), 1437–1480.
26. Gates, B. Trustworthy computing. Microsoft memo (Jan. 15, 2002); [wired.com](https://www.microsoft.com/en-us/research/wp-content/uploads/2002/01/trustworthy-computing.pdf)
27. Gu, R., Shao, Z., Chen, H., Wu, X.N., Kim, J., Sjberg, V. and Costanzo, D. Certikos: An extensible architecture for building certified concurrent OS kernels. *Proceedings of 12th USENIX Symp. Operating Systems Design and Implementation*, 2016.
28. Gulwani, S., Polozov, O. and Singh, R. Program Synthesis. *Foundations and Trends® in Programming Languages*. Now Publishers Inc., 2017.
29. Harrison, J. Formal verification at Intel. In *Proceedings of the 18th Annual IEEE Symp. Logic in Computer Science*. IEEE, 2003.
30. Hawblitzel, C., Howell, J., Lorch, J.R., Narayan, A., Parno, B., Zhang, D. and Zill, B. Ironclad apps: End-to-end security via automated full-system verification. In *Proceedings of the 11th USENIX Symp. Operating Systems Design and Implementation*, 2014.
31. Hutchinson, B., Mitchell, M., Xu, C., Doshi, T. Fairness indicators: Thinking about fairness evaluation, 2020; https://www.tensorflow.org/tfx/fairness_indicators/guidance.
32. Ivanov, R., Weimer, J., Alur, R., Pappas, G.J. and Lee, I. Case study: Verifying the safety of an autonomous racing car with a neural network controller. In *Proceedings of the 23rd ACM Intern. Conf. Hybrid Systems: Computation and Control*, 2020.
33. Kleinberg, J., Mullainathan, S. Raghavan, M. Inherent trade-offs in the fair determination of risk scores. In *Proceedings of Innovations in Theoretical Computer Science*, 2017.
34. Koh, N., Li, Y., Li, Y., Xia, L., Beringer, L., Honore, W., Minsky, W., Pierce, B.C. and Zdancewic, S. From C to interaction trees: Specifying, verifying, and testing a networked server. In *Proceedings of the 8th ACM SIGPLAN Intern. Conf. Certified Programs and Proofs*, Jan. 2019.
35. Kwiatkowska, M., Norman, G. and Parker, D. PRISM: Probabilistic Symbolic Model Checker. In *Proceedings of the PAM/PROBMIV'01 Tools Session*. Sept. 2001, 7–12. Available as Technical Report 760/2001, University of Dortmund.
36. Lamport, L. Proving the correctness of multiprocess programs. *IEEE Trans. Software Engineering* SE-3, 2 (Mar. 1977), 125–143; doi: 10.1109/TSE.1977.229904.
37. Mandal, D., Deng, S., Hsu, D., Jana, S. and Wing, J.M. Ensuring fairness beyond the training data. In *Proceedings of the 34th Conf. Neural Information Processing Systems*, 2020.
38. Manna, Z. and Pnueli, A. Verification of concurrent programs: Temporal proof Principles. *Workshop on Logic of Programs*. Springer-Verlag, 1981, 200–252.
39. Microsoft Azure blog. Fairness in machine learning models, 2020; <https://docs.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml>
40. Narayanan, A. 21 Definitions of fairness and their politics. In *Proceedings of FAT* 2018*. Tutorial; <https://www.youtube.com/watch?v=jIXIuYdnyk>.
41. National Research Council. *Trust in Cyberspace*. The National Academies Press, 1999; <https://doi.org/10.17226/6161>
42. National Science Foundation. National AI Institutes Call for Proposals, 2019; <https://www.nsf.gov/pubs/2020/nsf20503/nsf20503.htm>.
43. National Science Foundation. NSF Program on Fairness in Artificial Intelligence in Collaboration with Amazon (FAI), 2020; https://www.nsf.gov/funding/pgm_summ.jsp?pins_id=505651
44. Newcombe, C., Rath, T., Zhang, F., Munteanu, B., Brooker, M. and Dearndeuff, M. How Amazon Web Services uses formal methods. *Commun. ACM* 58, 4 (Apr. 2015), 66–73.
45. Networking and Information Technology Research and Development Subcommittee, Machine Learning and Artificial Intelligence Subcommittee, and the Special Cyber Operations Research and Engineering Subcommittee of the National Science and Technology Council. Artificial Intelligence and Cybersecurity: Opportunities and Challenges. Public Report; <https://www.nitrd.gov/pubs/AI-CS-Tech-Summary-2020.pdf>.
46. Platzer, A. *Logical Foundations of Cyber-Physical Systems*. Springer, Cham, 2018.
47. Pnueli, P. The temporal logic of programs. In *Proceedings of the Symp. Foundations of Computer Science*, 1977, 46–57.
48. Pong, F. and Dubois, M. Verification techniques for cache coherence protocols. *ACM Computing Surveys* 29, 1 (Mar. 1997).
49. Protzenko, J. et al. Verified low-level programming embedded in F*. In *Proceedings of 22nd Intern. Conf. Functional Programming*, May 2017.
50. Queille, J.P. and Sifakis, J. Specification and verification of concurrent systems in CESAR. In *Proceedings of the Intern. Symp. Programming*, LNCS 137, 1982, 337–351.
51. Seshia, S.A. et al. Formal specification for deep neural networks. In *Proceedings of the Intern. Symp. Automated Technology for Verification and Analysis*, LNCS 11138, Sept. 2018.
52. Shankland, S. Facebook starts building AI with an ethical compass. CNET, 2018; <https://www.cnet.com/news/facebook-starts-building-ai-with-an-ethical-compass/>.
53. Thomas, P.S., da Silva, B.C., Barto, A.G., Giguere, S., Brun, Y. and Brunskill, E. Preventing undesirable behavior of intelligent machines. *Science* 366, 6468 (2019), 999–1004.
54. Tiwari, P. et al. Computer-extracted texture features to distinguish cerebral radionecrosis from recurrent brain tumors on multiparametric MRI: A feasibility study. *American J. Neuroradiology* 37, 12 (2016), 2231–2236.
55. Varshney, K. Introducing AI Fairness 360. IBM, 2018; <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/>.
56. Wang, S., Pei, K., Whitehouse, J., Yang, J. and Jana, S. Formal security analysis of neural networks using symbolic intervals. In *Proceedings of the 27th USENIX Security Symp.*, 2018.
57. Wang, S., Pei, K., Whitehouse, J., Yang, J. and Jana, S. Efficient formal safety analysis of neural networks. In *Proceedings of Neural Information Processing Systems*, 2018.
58. White House. Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, Dec. 3, 2020; <https://www.whitehouse.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/>.
59. Younes, H.L.S. and Simmons, R.G. Probabilistic verification of discrete event systems using acceptance sampling. In *Proceedings of the 14th Intern. Conf. Computer Aided Verification*, (Copenhagen, Denmark, July 2002), E. Brinksma and K. Guldstrand Larsen, Eds. Lecture Notes in Computer Science 2404, 223–235.
60. Zheng, D. and Kifer, D. Light DP: Towards automating differential privacy proofs. In *Proceedings of the 44th ACM SIGPLAN Symp. Principles of Programming Languages*, 2017, 888–901.

Jeannette M. Wing (wing@columbia.edu) is Avansians Director of the Data Science Institute and Professor of Computer Science at Columbia University, New York, NY, USA.

© 2021 ACM 0001-0782/21/10



Watch the author discuss this work in the exclusive *Communications* video. <https://cacm.acm.org/videos/trustworthy-ai>

Introducing *ACM Transactions on Human-Robot Interaction*

Now accepting submissions to ACM THRI

In January 2018, the *Journal of Human-Robot Interaction* (JHRI) became an ACM publication and was rebranded as the *ACM Transactions on Human-Robot Interaction* (THRI). It will continue to be open access, fostering the widest possible readership of HRI research and information. All issues will be available in the ACM Digital Library.

ACM THRI aims to be the leading peer-reviewed interdisciplinary journal of human-robot interaction. Publication preference is given to articles that contribute to the state of the art or advance general knowledge, have broad interest, and are written to be intelligible to a wide range of audiences. Submitted articles must achieve a high standard of scholarship. Accepted papers must: (1) advance understanding in the field of human-robot interaction, (2) add state-of-the-art or general information to this field, or (3) challenge existing understandings in this area of research.

ACM THRI encourages submission of well-written papers from all fields, including robotics, computer science, engineering, design, and the behavioral and social sciences. Published scholarly papers can address topics including how people interact with robots and robotic technologies, how to improve these interactions and make new kinds of interaction possible, and the effects of such interactions on organizations or society. The editors are also interested in receiving proposals for special issues on particular technical problems or that leverage research in HRI to advance other areas such as social computing, consumer behavior, health, and education.

The inaugural issue of the rebranded *ACM Transactions on Human-Robot Interaction* has been published and can be found in the ACM Digital Library.

For further information and to submit your manuscript visit thri.acm.org.



Association for
Computing Machinery

research highlights

P. 74

**Technical
Perspective
Liquid Testing
Using Built-in
Phone Sensors**

By Tam Vu

P. 75

Liquid Testing with Your Smartphone

By Shichao Yue and Dina Katab

P. 84

**Technical
Perspective
The Real-World
Dilemma of Security
and Privacy by
Design**

By Ahmad-Reza Sadeghi

P. 85

Securing the Wireless Emergency Alerts System

By Jihoon Lee, Gyuhong Lee, Jinsung Lee, Youngbin Im, Max Hollingsworth, Eric Wustrow, Dirk Grunwald, and Sangtae Ha

Technical Perspective

Liquid Testing Using Built-in Phone Sensors

By Tam Vu

HOW MANY TIMES have you wondered if the tap water in the hotel you just checked into is safe to drink? Have you ever wanted to check the amount of alcohol concentration in a drink served by a kind bartender? If you have, you might be able to find the answers right from your mobile phone using the new technique recently introduced by a group of MIT computer scientists to measure surface tension of a liquid. This technology can potentially not only confirm any water contamination and measure alcohol concentration, but also identify the presence of substances in liquid for diagnosis in healthcare or detect counterfeit luxury goods such as brandy or perfume.

Existing liquid testing methods are often based on a liquid's measurable properties such as electric permittivity, optical absorption, and so on. Surface tension is another property of liquids, representing the tendency of liquid surfaces to shrink into the minimum surface area possible, resulting from the greater attraction of liquid molecules to each other. However, it is one of the most difficult properties to measure and requires high-cost and sophisticated instruments, called tensiometers. Examples of these tensiometers include Du Noüy ring, Wilhelmy plate, spinning drop, pendant drop, bubble pressure, and acoustic levitation. Although such tensiometers can achieve high accuracy, their measurements are often performed in-lab due to the complexity of hardware, preparation, and manual procedures. To address these drawbacks, recent efforts on measuring surface tension using mobile devices have shown some promising results. However, its current poor performance and complexity are the two main reasons for its low adoption.

The following paper proposes a novel technique for determining the surface tension of a liquid by leveraging the optical absorption of waves


propagating on the fluid surface. Using the equation for surface tension, this method tries to infer the wavelength of capillary waves—the key relation to the given liquid property. However, making a correct inference of the wavelength is challenging. Specifically, the first challenge stems from very shallow capillary waves produced by the phone vibrations on the liquid surface inside the container. Another challenge is due to the basic mechanism of the camera-rolling shutter, which leads to unexpected motion artifacts. When these challenges combine, it not only blurs the wave pattern in the photos captured by the phone's camera but also complicates the processing of these photos for wavelength estimation.

Without adding supportive hardware, the authors propose a neat set of techniques to achieve equivalent performance to in-lab manual tensiometers that confidently demonstrate the success of this research. One remarkable insight that helps the authors arrive at their solutions is the observation that the reflection of the shallow

The following paper proposes a novel technique for determining the surface tension of a liquid by leveraging the optical absorption of waves propagating on the fluid's surface.

waves at the bottom of the container by the phone's flashlight correlate to the clearness of their pattern in the photos, which helps solve the first challenge. In addition, the authors leverage the intrinsic working mechanism of the hardware to pinpoint the motion artifacts sourcing from the camera-rolling shutter, which helps resolve the second challenge. As a result, their proposed technique can reduce the inherent poor performance of wavelength inference in existing approaches.

Applying the proposed solutions, the implemented system is able to gain very intriguing results, which have an absolute surface tension error of 0.75mN/m, compared to an in-lab manual tensiometer with a resolution of 0.5mN/m. Interestingly, the proposed method can measure the surface tension with values very close to the actual measurement in alcohol concentration, water contamination detection, and protein-level tracking in urine—a pivotal physiological index used in controlling diabetes and kidney disease.

These results illustrate the proposed technique is the first thorough mobile application able to precisely measure liquid surface tension simply using the camera, the flashlight, and the built-in vibro-motor on smartphones. This line of research opens multiple potential applications as well as research directions. The very next logical steps along this line of research includes improving its performance, applying the technique to other practical applications such as food safety, counterfeit goods detection, and validating such fakery in real-world settings with real-world deployments. 

Tam Vu is an associate professor at the University of Colorado, Boulder, CO, USA, where he founded and directs the Mobile and Networked Systems (MNS) Lab.

Copyright held by author.

Liquid Testing with Your Smartphone

By Shichao Yue and Dina Katabi

Abstract

Surface tension is an important property of liquids. It has diverse uses such as testing water contamination, measuring alcohol concentration in drinks, and identifying the presence of protein in urine to detect the onset of kidney failure. Today, measurements of surface tension are done in a lab environment using costly instruments, making it hard to leverage this property in ubiquitous applications. In contrast, we show how to measure surface tension using only a smartphone. We introduce a new algorithm that uses the small waves on the liquid surface as a series of lenses that focus light and generate a characteristic pattern. We then use the phone camera to capture this pattern and measure the surface tension. Our approach is simple, accurate, and available to anyone with a smartphone. Empirical evaluations show that our mobile app can detect water contamination and measure alcohol concentration. Furthermore, it can track protein concentration in the urine, providing an initial at-home test for proteinuria, a dangerous complication that can lead to kidney failure.

1. INTRODUCTION

Mobile computing has recently seen a surge in research on inexpensive methods for measuring liquid properties and identifying liquid type.^{7, 12, 17, 21} The developed methods can detect water contamination and distinguish a variety of liquid types such as water, milk, oil, and different alcohol concentrations. The goal of this line of research is to enable liquid testing outside the lab environment and encourage ubiquitous applications. However, the proposed designs require a specialized setup (a robot,²¹ a special container,⁷ etc.) and use devices typically unavailable to the general population (e.g., UWB radios or RFID readers). Although they make an important step toward ubiquitous liquid testing, they are still difficult to use by lay users.

This paper asks whether it is possible to deliver such services to lay users without a specialized setup, and using *only* a device that almost everyone has: a smartphone. Answering this question is not simple. Typically, liquid testing is done by measuring a particular property, such as electric permittivity or optical absorption, and using the measurements to identify the liquid type and characteristics.^{7, 12, 17, 21} However, none of the properties used in past work can be measured with a smartphone. To address this problem, we explore an alternative liquid property, surface tension, and develop algorithms and system architectures that enable measuring surface tension using only a smartphone.

Surface tension characterizes the force that holds the surface molecules together and minimizes the surface area. Measuring surface tension can reveal water

contamination and allow for distinguishing liquid types.¹⁹ Water has a relatively high surface tension, and when polluted with organic compounds such as bacteria, oil, petroleum or its derivatives, its surface tension decreases significantly.^{3, 19} Hence, one may use this property to detect water contamination. Further, lipids and proteins act as surfactants, that is, they reduce surface tension. Alcohol also reduces surface tension, a property that can be leveraged to measure alcohol concentration.²⁰ Most interestingly, the ability to measure surface tension at home can enable early detection of medical problems. For example, low surface tension of urine may indicate the presence of excess protein, a dangerous complication in diabetes patients.⁸ Daily measurements of urine surface tension help detect diabetic nephropathy (the chronic loss of kidney functions) and monitor the effect of treatment.⁸

Today, measurements of surface tension require a device called tensiometer, which typically costs thousands of dollars.⁹ They are often conducted by dipping a platinum plate into the liquid and carefully measuring the force required to pull it out. The process is complicated and requires professional training.⁹ This high bar hampers the ubiquitous application of surface tension.

We introduce CapCam (**Capillary Camera**), the first mobile app that measures liquid surface tension using only a smartphone. To measure surface tension, the user places the smartphone on top of a lightweight container, such as a paper cup, and activates the app, as shown in Figure 1. The phone vibrates and forces the container's wall to vibrate. The vibration generates capillary waves on the liquid surface, that is, small waves whose wavelength characterizes the liquid's surface tension. Our app uses the flashlight camera to take a few photos of the liquid, which it processes to estimate the capillary wavelength and hence the surface tension. Our approach is accurate, simple, cheap, and accessible to any user with a smartphone.

Measuring capillary waves just using smartphones is quite challenging for two reasons. First, the waves are very shallow¹; simply trying to image them at the liquid surface does not yield accurate results. Thus, instead of imaging the capillary waves directly, we image their reflections at the bottom of the container. We model the small waves on the surface as a series of convex and concave lenses. When illuminated with the flashlight on the phone, the lenses focus the light and create a pattern of bright and dark rings

The original version of this paper was published in *Proceedings of ACM MobiSys '19*.

Figure 1. CapCam setup: the phone is placed on a paper cup. Capillary waves are generated by the vibro-motor inside the smartphone. Then with the flashlight, the camera of the phone can capture a bright-and-dark pattern, from which we can calculate the surface tension of the liquid.



on the bottom of the container. The focusing effect of the small lenses makes this pattern much higher contrast than the waves on the surface, and hence easier to capture and characterize using a camera.

The second challenge stems from the camera rolling shutter. Phone cameras do not capture images in one snapshot; they capture a picture by sequentially scanning rows of photo diodes.¹⁴ As the waves are moving, the sequentially scanning creates motion artifacts in the captured image. If uncorrected, these artifacts can lead to large measurement errors. CapCam addresses this problem by recognizing that the direction orthogonal to the scanning direction is not affected by motion artifacts; it processes the image to identify this orthogonal direction and uses the line of pixels along that direction to measure capillary waves.

We have implemented CapCam on an iPhone X and evaluated its performance by comparing its output against that of a high-end digital tensiometer. Our evaluation shows that CapCam accurately measures liquid surface tension using only a smartphone. Specifically, CapCam has an averaged absolute surface tension error of only 0.75 mN/m. In comparison, an entry-level manual tensiometer¹³ has a resolution of 0.5 mN/m, costs several thousand dollars, and requires expert knowledge. CapCam has enough resolution to detect small differences in alcohol concentration of 0.5%, whereas past work is limited to large differences of 20% for Nutrilizer¹⁷ and 25% for RFIQ.¹² CapCam's resolution allows it to easily detect water contamination and it also has enough sensitivity to track changes in protein levels in urine as they transition from healthy to dangerous levels, providing an initial at-home test of proteinuria.

Contributions: this paper makes the following contributions:

1. It introduces CapCam, the first design that estimates liquid surface tension using only a smartphone, without any specialized hardware. CapCam is also the first mobile app that detects water contamination, measures alcohol concentration, and tracks changes in protein levels in urine.

2. It presents two novel algorithms: (1) an algorithm for inferring capillary wavelength by creating a characteristic pattern on the bottom of the container, and (2) an algorithm for characterizing the impact of rolling shutter on capillary waves and eliminating the resulting measurement errors.
3. It provides an implementation and empirical evaluation that demonstrate the efficacy of the proposed design.

2. RELATED WORK

2.1. Liquid testing in mobile and ubiquitous computing

The topic of liquid testing has recently attracted significant interest.^{7,12,21,23} Most proposals try to infer electric permittivity,^{7,21,23} a property that characterizes how a liquid affects radio waves. For example, TagScan²¹ and LiquID⁷ measure the time delay and power attenuation incurred by an RF signal as it traverses the liquid of interest. Their approach requires a complex setup (a moving robot, or a particular container) and uses special radios not typically used by lay users. Further, as RF attenuation and phase are highly sensitive to liquid depth, these systems have to be carefully calibrated. As a result, these methods either exhibit relatively large errors (10% in LiquID⁷) or they avoid measuring exact values and resort to classifying the liquid as one of a few known types (as in TagScan²¹). There are also proposals that identify liquids using RF coupling, as in RFIQ,¹² or RF reflections as in RadarCat.²³ They too use special radios (Soli or RFID readers); further they rely on a classifier to distinguish a few liquid types or a few concentration levels and do not generalize to unseen liquid types or concentration levels.

Some proposals rely on optical absorption. Specifically, when shining intensity-modulated light on a liquid, different liquids produce uniquely different sound spectra. These solutions require custom hardware and have a relatively limited resolution. For example, by analyzing the received spectra, Nutrilizer¹⁷ can predict alcohol concentration level with a limited resolution (20%).

CapCam is inspired by the above work, but focuses on liquid testing using a smartphone, a device that almost every user has.

2.2. Measuring surface tension

Today, surface tension is measured in the lab using an expensive device called tensiometer. There are four common types of tensiometers: force-based (e.g., Wilhelmy plate), pendant-drop-based, contact-angle-based, and capillary-waves-based. Although tensiometers can achieve high accuracy, expensive instruments and sophisticated measurement procedures prevent their ubiquitous use. In comparison, CapCam extends the model based on capillary waves to allow a nonexpert user to measure surface tension using only a smartphone.

Although some past work has attempted to measure surface tension using a smartphone, all past proposals require additional complex hardware, and none is able to complete the measurements with only a smartphone.^{4,5,11,22} Specifically, the work in Wei et al.²² uses a cellphone camera to capture images of capillary waves. However, it ignores the rolling shutter effect, which leads to poor performance.

Further, it requires custom hardware where a large container is elevated between two stands although projecting light from the bottom, a signal generator that excites the liquid, and a paper screen held on top of the surface. The proposals in Goy et al.¹¹ and Chen et al.⁵ use cellphone cameras to capture images of a drop of liquid. Their method requires specialized equipment to exercise tight control of the drop's size and shape. It also requires extreme cleanliness and a complex measurement procedure making it hard to conduct by nonexperts. Similarly, the method in Chen et al.⁴ requires custom hardware to control the size of the liquid drop as it touches a special solid surface.

CapCam builds on the above work. However, unlike these methods, CapCam can measure surface tension simply using a smartphone app without any special hardware. Further, it can be operated by an average user without any training.

3. CAPILLARY WAVES PRIMER

In this section, we provide a primer on capillary waves and their relation to surface tension. For waves propagating on a fluid surface, *gravity* (g) and *surface tension* (γ) are the forces that make the liquid restore at surface. Those forces control the dispersion relation of the wave, which relates its *wavelength* (λ) to the vibration *frequency* (f) and liquid *density* (ρ), and can be expressed by the following equation (see Chapter 16 of Blandford² for a detailed derivation):

$$(2\pi f)^2 = g(2\pi/\lambda) + \frac{\gamma}{\rho}(2\pi/\lambda)^3. \quad (1)$$

When the waves are large, such as those observed on the surface of a lake, they are dominated by the gravity term g and referred to as gravity waves. When the waves are small, the effect of gravity is negligible, and the waves are dominated by surface tension γ ; they are called capillary waves.

We can re-order the terms in the above equation to measure surface tension as follows:

$$\gamma = \frac{(2\pi f)^2 - g(2\pi/\lambda)}{(2\pi/\lambda)^3} \rho. \quad (2)$$

This equation provides a procedure for measuring surface tension using capillary waves. Specifically, we can use a vibration source to generate capillary waves on the liquid's surface. Knowing the vibration frequency f , we can substitute the gravity term and the liquid density from the corresponding data sheets. Hence, all we need is to measure the wavelength λ in order to measure the surface tension.

The challenge however is that capillary waves are very shallow, that is, the displacement they cause in the liquid surface is one to a few microns. This is about one tenth of the average thickness of a human hair. Thus, such waves are invisible and their measurement typically requires dedicated laboratory equipment²⁴ with a complicated setup and procedure. In the rest of this paper, we describe CapCam, a novel design that can measure capillary wavelength using only a smartphone.

4. CAPCAM DESIGN

CapCam measures surface tension using a phone's camera, flashlight, and vibro-motor. The process is very simple. The user places the smartphone on top of the container (e.g., a cup) and activates the CapCam app, as shown in Figure 1. CapCam uses the phone's vibro-motor to generate capillary waves on the liquid surface and then takes a few photos of the vibration pattern using the phone flashlight camera. It uses a series of algorithms to process these photos to infer the capillary wavelength and hence the surface tension.

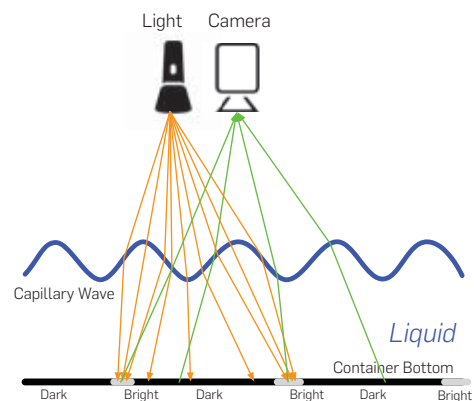
CapCam's inference algorithm addresses the challenge of measuring very shallow capillary waves and makes it possible to measure the wavelength with only a smartphone. In this section, we describe the model and analysis underlying CapCam's wavelength inference algorithm. We describe how we deal with phone hardware in the next section.

4.1. CapCam's wavelength inference algorithm

We model the crests and troughs of capillary waves as a series of convex and concave lenses. By shining a flashlight on the waves from the top, we can create a visible pattern that reflects off the bottom of the container. Specifically and as shown in Figure 2, when the light goes through a wave crest, it is focused into a small area on the bottom of the container. On the other hand, when the light goes through a wave trough, it diverges causing a dark region on the bottom of the container. This results in a series of bright and dark rings on the bottom of the container.

If we try to capture this pattern using a smartphone, the camera would see only the reflections that exit the liquid surface and propagate through air toward the camera. However, due to reciprocity, light reflected from the bottom toward the surface experiences the opposite effects of light entering the surface from air, that is, light rays from

Figure 2. An cross-section illustration of the setup. Because of refraction, light rays are focused/diverged by the crests/troughs of the waves, resulting in a series of dark and bright rings on the bottom.



Liquid density is a constant given temperature and pressure. For most applications, one can assume room temperature and atmosphere pressure. If the measurements are conducted at unusually high/low temperatures or very high elevation, one should substitute the corresponding liquid density from the liquid's data sheets.

the bright rings will diverge and light rays from the dark rings will converge. Hence, if the camera is exactly at the same position as the flashlight, it will not see any pattern. However, the camera is never at the exact same location as the flashlight, and the light rays that it receives do not trace the exact path of the incoming light. As a result, the focus and divergence effects that the light experienced although traversing the surface from air to liquid do not get cancelled as it traverses the surface from liquid to air. Therefore, the pattern continues to be visible to the camera albeit at lower contrast.

After we take an image of the pattern at the bottom of the container, we measure the distance between two consecutive bright rings in terms of pixels, denoted as p . We need to convert p from pixel-based distance to real distance. Of course, this depends on the distance between the camera and the bottom of the container, which we denote as $d + h$, where d is the distance between the camera and the surface of the liquid, and h is the depth of the liquid. Let us use r_{d+h} to refer to the resolution of the camera for objects at distance $d + h$. Then, we can convert the distance between consecutive bright rings in the image, p , into real distance, λ_b , as follows:

$$\lambda_b = p / r_{d+h} \quad (3)$$

Next, we want to use the interval between two bright rings to estimate the capillary wavelength. To calculate the relationship between the two, we approximate the surface waves as perfect lenses; when the incident ray goes through the center of a lens, its direction is unchanged, as shown in Figure 3. Thus, the relation between the capillary wavelength λ and the interval between two bright rings, λ_b , is

$$\lambda = (d / d + h) \cdot \lambda_b \quad (4)$$

Combining the two equations, we have

$$\lambda = \frac{d}{d+h} \cdot \lambda_b = \frac{d}{d+h} \cdot \frac{p}{r_{d+h}} \quad (5)$$

As camera resolution is inversely proportional to distance from the imaged object, we have

$$r_1 * d_1 = r_2 * d_2 \quad (6)$$

Therefore, we can rewrite Equation (5) as follows:

$$\lambda = \frac{p}{r_{d+h} * (d+h) / d} = \frac{p}{r_d} \quad (7)$$

where p is the number of pixels between two consecutive bright rings, and r_d is the camera resolution for objects at distance d , that is, at the liquid surface.

Equation (7) and the model underlying it provide us with an algorithm to compute the capillary wavelength. They also show that we only need to measure the distance between the camera and the liquid surface and use the corresponding resolution to convert the inter-ring pixels in the image to the actual capillary wavelength.

Note that, it is enough to calibrate the camera resolution at one default distance based on Equation (6). The resolution r at distance d can be computed as

$$r = (r_0 * d_0) / d \quad (8)$$

where r_0 is the resolution at the default distance d_0 .

4.2. Illustrative simulation

To provide a better visual understanding of CapCam's algorithm, we build a simulator based on ray tracing. The surface of the liquid and the bottom of the container are discretized into dense pixels. Our simulation can be divided into two phases. First, the light source emits light rays onto each pixel on the surface, and for each light ray, we calculate its location of arrival at the bottom of the container. Then the intensity of each bottom pixel is set to the total number of rays falling on that pixel. In the second phase, we densely sample locations inside each pixel in the camera and trace the light rays that fall on the camera to their origin on the bottom of the container. This allows us to obtain the set of pixels on the bottom of the container that are reachable from that camera pixel. The intensity at each camera pixel is set to the sum of the intensity of all reachable bottom pixels.

We show the results of a simulated experiment in Figure 4. By comparing Figure 4d with Figure 4e, we can see that the peaks of the pattern on the bottom of the container are much sharper than the crests of the capillary waves at the surface. This is because the crests of the surface waves focus light into a series of much narrower bright rings at the bottom of the container, therefore increasing the contrast. Also by comparing Figure 4f with Figure 4e, we can see that troughs of the pattern captured by the camera are distorted and the peaks are shifted to the left. This is because the slight separation between the camera and the light source which causes the camera to see the pattern from its own angle. But still, we can use the imaged pattern to calculate the wavelength accurately using Equation (7). In this simulation, the camera has a resolution of 40 pixel-per-millimeter at the surface distance, and the wavelength is 3 mm. At these values, and assuming the liquid is water, the wavelength is exactly 120 pixels.

Figure 3. An illustration of the wavelength calculation. We plot only the rays that pass through the center of the crests of the waves, which do not bend.

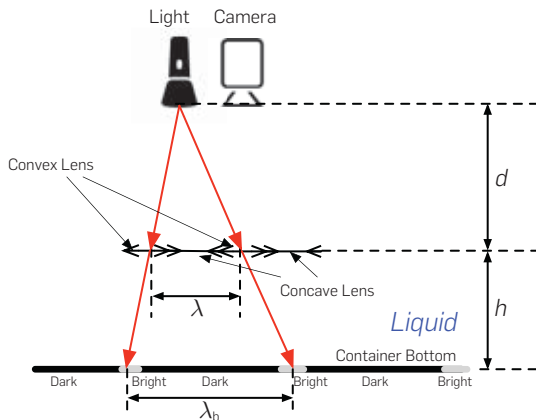
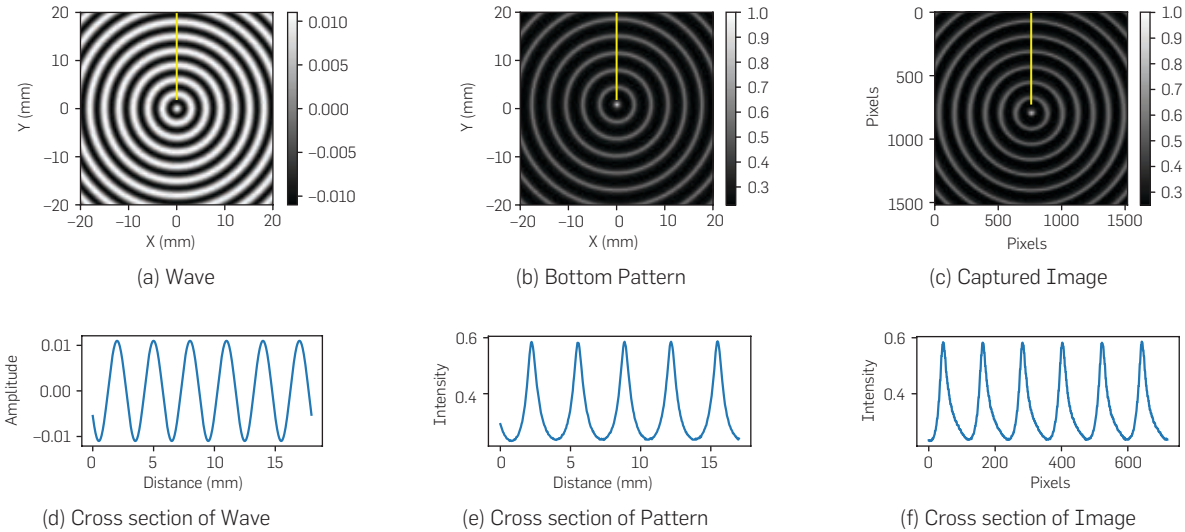


Figure 4. A simulated experiment showing the relation between the waves on the surface, the pattern on the bottom of the container, and the captured image. Figures on the second row show the pixel value along the yellow line in the corresponding first row figure. Comparing (d) with (e), we see that the crests of the pattern on the bottom are both sharper and higher value than the crests of the waves on the surface. This is due to the focusing effect on the lenses. Also note that in (f), the crests are shifted to the left; this is because the camera and the light are not co-located.



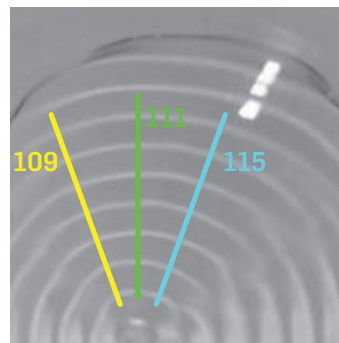
5. DEALING WITH THE CAMERA'S ROLLING SHUTTER

The algorithm in Section 4 measures the surface tension using the following: a high frequency vibro-motor that can be used to excite capillary waves, a light source to provide illumination, and a camera to capture the resulting pattern. Thankfully, a smartphone includes all three components. Thus, in principle, we can measure the wavelength in the image captured by the phone and convert from pixel-based distance into real distance using Equation (7), which gives us the surface tension. However, in practice, we still need to take care of another challenge caused by the reality of phone hardware: the camera's rolling shutter effect.

As explained earlier, the wavelength is the distance between two consecutive bright rings. Because the rings are concentrated (see Figure 4c), we can pick a particular direction along the radius of the rings and use it to measure the distance between consecutive bright rings, which would then yield the wavelength. Unfortunately, it is not that simple. In reality, not all directions give the correct wavelength. Figure 5 shows an example image of the ring pattern due to capillary waves. The figure shows three radial directions along and the corresponding wavelength in number of pixels. The figure shows that the average number of pixels between two consecutive bright rings along these three directions is 109, 111, and 115. The differences between these estimates of the wavelength cause a significant difference in the resulting estimate of surface tension. In fact, a difference of two pixels leads to an error of 4 mN/m in surface tension, which is more than 5% of the water's surface tension $\gamma = 72 \text{ mN/m}$. This means that we have to figure out which of these radial directions yield the correct estimate of wavelength and use only that direction.

In order to pick the correct radial direction for our

Figure 5. A sample image taken by a smartphone. The three lines represent three radii with different directions. Numbers besides the lines are the corresponding wavelength measured in number of pixels, and they are different from each other. This suggests picking wrong radii may result in a significant error.



estimation, we first need to understand why different directions yield different distances between the bright rings, that is, different wavelengths. The reason is the *rolling shutter*. Rolling shutter refers to that a camera captures a picture by sequentially scanning rows of photo diodes.¹⁴ Therefore, pixels are not recorded at exactly the same instant, and because the wave is traveling rapidly, this will result in a *Doppler effect*.

In the example in Figure 5, the camera is scanning from right to left, and the wave is propagating from the edge of the cup to the center. The propagation direction and the scanning direction are aligned for the radius on the right (in the figure) and are opposite for the radius on the left. Thus, the measured wavelength is shorter along the left radius and longer along the right radius. As for the radius in the middle,

the propagation direction of the wave is perpendicular to the camera scanning direction, and hence the measurement is *not* affected by the rolling shutter.

Therefore, for accurate estimation, we want to measure the wavelength along the radius perpendicular to the camera's scanning direction (the green line in Figure 5). Given that the camera scans images horizontally, the correct radius is along the vertical direction in the image. We can ask the user to manually pick the vertical radius. However, it is preferable to do it automatically to reduce user overhead and any potential errors. To do so, we search for the vertical line in the image that maximizes symmetry. Specifically, because the line we are looking for is a radius of the concentrated rings, the pattern on its left should be similar to the pattern on the right (i.e., the intensity of the few pixels to the left of the line is similar to the intensity of the few pixels to the right of the line). Thus, we only need to find a vertical line whose neighbor pixels are symmetric with respect to itself. This is similar to detecting reflection symmetry in images and can be solved by convolving with a wavelet filter.⁶ Specifically, we design a convolution filter that is -1 on the left side and $+1$ on the right side. When convoluting with an image, result will be close to 0 only when there is reflectional symmetry in the reception field of the filter.

The reflection symmetry is only observable when the reception field of the filter is large enough, for example, at the scale of one wavelength, which is about a hundred pixels. To increase the reception field of the filter, instead of using a big filter which is expensive, we can down-scale the image to speed up the convolution process. Then for each column in the convolved image, we calculate the sum of the absolute values of all of its pixels. Because of the property of the filter, the result in the ideal column should have the lowest value. Therefore, we select the column with minimum sum as the ideal location to measure the wavelength. Note that, in the implementation, we also average across multiple images to improve robustness of the algorithm.

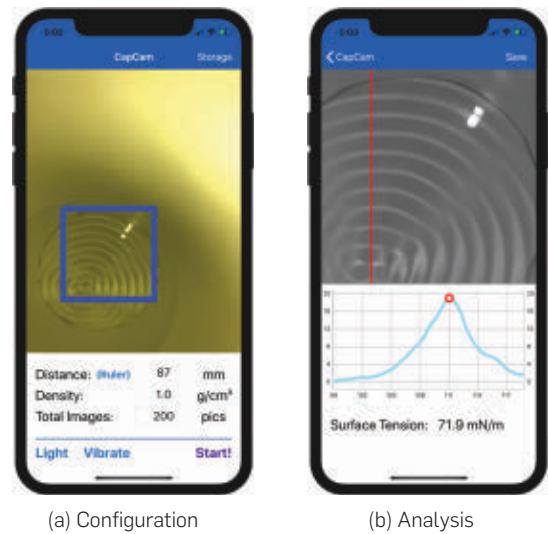
This algorithm can locate the vertical radius accurately. Further, because it is based on very simple operations, it is highly efficient and can run on an ordinary smartphone.

6. USER INTERFACE AND IMPLEMENTATION

We implement CapCam as a standalone iOS App using the Swift programming language. To speedup processing, we use Apple's Accelerate Framework for SIMD operations such as convolution and summation. This yields a significant speedup compared to a naive Swift implementation.

Two main user interfaces of CapCam are as shown in Figure 6. The UI on the left (Figure 6a) is the configuration interface. On this interface, we introduce each component from top to the bottom. The first component is the preview of the camera. On the preview, there is a blue frame showing the area where we run our analysis algorithm. By cropping the image, we are able to achieve a much faster speed without sacrificing accuracy. Next there are three text fields. The first text field is for the user to input the distance between the surface and the camera. The app includes a ruler to assist the user in measuring the distance directly with the

Figure 6. CapCam's user interface.



phone. The second text field is for the density, and its default value is 1.0. The third text field is for the user to specify the number of captured images. Having more images will increase the accuracy, yet will take a longer time. Then there is a progress bar indicating the progress of the analysis process. Finally, there are three buttons on the bottom. "Light" and "Vibrate" are the controls for the flashlight and the vibro-motor, and the "Start!" button is for starting the analysis process.

After the analysis finishes, the app shows the analysis interface, which is the UI on the right (Figure 6b). One sample captured image is shown on the top half, with the automatically detected estimation radius highlighted. The histogram of the wavelength is plotted beneath the sample image. Finally, the estimate of the surface tension is printed out at the bottom.

7. EVALUATION

In this section, we evaluate the performance of CapCam and its ability to deliver interesting applications to the user.

7.1. Ground truth

In all experiments, the ground truth of the surface tension is obtained by using an advanced digital force tensiometer (Dataphysics DCAT 11⁹), as shown in Figure 7. It measures the surface tension using the Wilhelmy plate method and can provide a resolution of 0.1 mN/m (0.1 millinewton per meter).

7.2. Experiment setup

We install CapCam on an iPhone X. Unless specified otherwise, we use a standard paper cup as a testing container and set the liquid depth to 45 mm. We place the phone on top of the cup, as shown in Figure 1. We hold the phone with our hand as shown in the figure to ensure it stays still while taking the images and does not move due to vibrations. The cup has a height of 132 mm. The camera on the iPhone is measured to have a resolution of 39.5 pixels per millimeter

at a distance of 87 mm, and the vibration frequency of the iPhone is centered at 144.5 Hz. For each measurement, CapCam continuously takes 200 images. On average, each surface tension measurement takes 8 s. For each liquid sample, we repeat the measurement five times and compute the average and standard deviation.

7.3. Detecting water contamination

Water has a relatively high surface tension, and when polluted with organic compounds such as petroleum, bacteria, pesticides, oil or its derivatives, water surface tension decreases significantly.¹⁹ In this section, we empirically evaluate the effectiveness of CapCam at detecting changes in surface tension due to such contamination. Although not all sources of water contamination change surface tension (e.g., metal contamination), our approach covers a large and important class of contaminants. Such bacterial and organic contamination is common anywhere unsanitary conditions are present. Further, people who live downstream from factories are at greater risk of contamination from petroleum and organic waste, and people in farming communities are at risk for contamination from agriculture waste.¹⁶

We compare five different water sources: (1) deionized (i.e., pure) water, (2) tap water, (3) rain water, (4) pond water from a tree pit, and (5) water left exposed for a week. Figure 8 plots the surface tension of the above water sources. The figure reveals two findings. First, by comparing the blue and red bars in the figure, we see that CapCam’s measurements of surface tension match those from the tensiometer. This shows CapCam’s accuracy. Second, the figure shows that both tap water and rain water have a surface tension similar to deionized water, which indicates that these sources of water are not polluted. On the other hand, pond water and exposed water have much lower surface tension, meaning

Figure 7. Digital tensiometer with a resolution of 0.1 mN/m for measuring ground truth of surface tension (Model: Dataphysics DCAT 11⁹).



A demo video of our experiments is available at [http:// people.csail.mit.edu/scyue/projects/capcam/](http://people.csail.mit.edu/scyue/projects/capcam/)

that they contain chemicals that can decrease surface tension, which is a sign of contamination.

7.4. Tracking protein in urine

People who have diabetes or high blood pressure are vulnerable to kidney disease. When the kidney is damaged, it starts leaking substrates into urine that are not supposed to be present. Many of these substrates reduce urine surface tension. In particular, microalbuminuria is a complication in diabetic patients, where the kidney starts leaking albumin into urine.¹⁵ We would like CapCam to help diabetic patients by providing them with an easy way to regularly test the level of albumin in urine. If the level increases beyond a safe threshold, the patient can contact the doctor for more intensive tests.

We empirically test CapCam’s effectiveness at delivering this application. We add different levels of egg albumin to a sample of healthy human urine and measure urine surface tension for different albumin concentrations.

Figure 9 shows urine surface tension as a function of albumin concentration. Note that a protein level of over 30 mg/L is the threshold at which the patient has microalbuminuria.¹⁰ The results in the figure indicate that our system can track changes in urine surface tension with

Figure 8. Water contamination detection. Both tap water and rain water have a surface tension close to deionized water. In contrast, pond water and exposed water have lower surface tension values due to contamination.

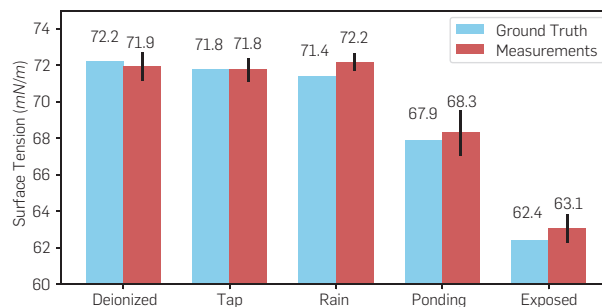
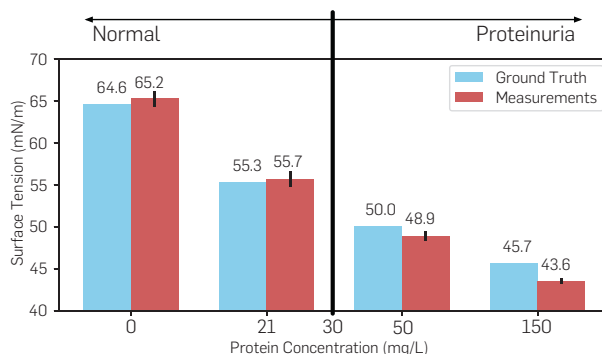


Figure 9. Urine surface tension test with different protein concentration levels. For a healthy person, protein concentration level in urine should be less than 30 mg/L,¹⁰ and the greater the concentration, the higher the risk.



increased albumin concentration in a manner comparable to a professional tensiometer. Further, it can detect when the protein concentration becomes dangerous. This means a patient can track the progress of the disease in the home using her smartphone.

7.5. Measuring alcohol concentration level

We test the performance of our system on a series of ethanol solutions with different concentration levels. This is a complex scenario because an ethanol solution is a mixture of two liquids: water and ethanol. The density of the mixture ρ in Equation (2) depends on the ratio of ethanol to water in the mixture. Thus, we are facing a chicken-and-egg problem: we want to measure surface tension to estimate alcohol concentration; however, we need the alcohol concentration to estimate ρ , which we need for our measurements of surface tension.

To address this issue, we leverage the fact that both the surface tension and the density of the ethanol mixture are functions of its concentration level. Further, they are known functions available in data sheets.^{18, 20} Given both functions, we can compute for each given ethanol concentration level, its surface tension and its density, and take the ratio of the two, denoted as the *surface tension to density ratio (TDR)*.

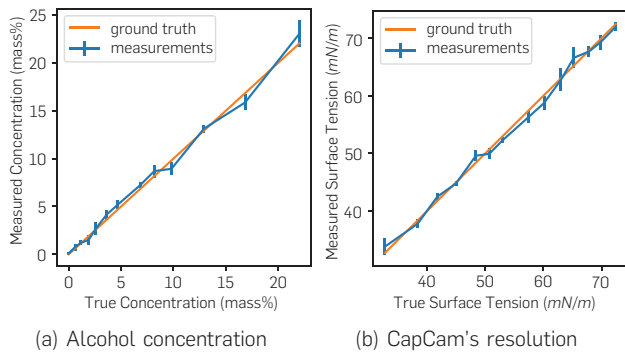
We can then rewrite Equation (2), so that we can estimate the TDR, that is, γ/ρ , as opposed to surface tension γ :

$$TDR = \frac{\lambda}{\rho} = \frac{(2\pi f)^2 - g(2\pi/\lambda)}{(2\pi/\lambda)^3}. \quad (9)$$

Notice that the left hand side of this equation is the TDR, and we can compute it by substituting for the vibration frequency, gravity, and the wavelength of the capillary waves, which we can compute as before. Therefore, we can first measure the TDR of the alcohol using CapCam, and then convert the TDR into the concentration level.

The measurement result for ethanol concentration level

Figure 10. (a) Alcohol concentration measured by CapCam in blue and ground truth in orange. (b) CapCam's resolution: differences between true and measured surface tension show that CapCam has an averaged error of 0.75 mN/m.



Egg albumin and human serum albumin both belong to albumin, a family of globular proteins, and sharing similar physical properties.

is plotted in Figure 10a. The absolute error is only 0.51%. This implies our system is accurate enough to distinguish German Riesling from Australian Riesling and Portuguese Rose from French rosés.

7.6. CapCam's resolution

Next, we are interested in understanding CapCam's resolution, that is, the expected error in its measurements of surface tension. To estimate this value, we leverage the results from the alcohol concentration experiment. Specifically, for any alcohol concentration level, there are datasheets that report the surface tension and density.²⁰ By substituting the density ρ in the TDR equation above, we compute CapCam's estimate of surface tension, γ , which we can compare against the true surface tension in the datasheet. The measurements are plotted in Figure 10b. On average, CapCam has a surface tension error of only 0.75 mN/m, for surface tension in the range from 33 mN/m to 72 mN/m. In comparison, an entry-level manual tensiometer¹³ has a resolution of 0.5 mN/m, and it still costs thousands of dollars and requires complicated procedure.

8. DISCUSSION AND LIMITATIONS

- (a) **Sensitivity:** as in past work,^{7, 21} CapCam measures a particular liquid property and uses the measurements to make certain inferences. However, for any of these systems, the inference has to be taken within the measurement context. For example, as CapCam measures surface tension, it is sensitive to bacterial and organic contaminants but cannot sense contamination by heavy metal because they do not change surface tension. Thus, when it infers contamination, the water is certainly impure but the inverse is not necessarily true. Similarly, if it detects two liquids to be different, then they are different (assuming no measurement error), but if it cannot differentiate them, they might still be different liquids that have the same surface tension.
- (b) **Container specification:** CapCam has certain requirements on the container type. First, the container should have a flat bottom. If the bottom is not at, there will be artifacts in the ring pattern, which affect the measurements. Second, the container should be relatively light so that it vibrates with the vibro-motor. Third, the current system is designed for circular containers. Waves are excited by the vibrating wall and propagate from the edge to the center. Based on the Huygens-Fresnel principle, when the wall is circular, the resulting waves are also circular, hence creating clear rings as described earlier. But if the container is not circular, the wave pattern will be much more complex. Addressing this scenario requires extending the model to account for interaction between waves that traveled different distances, which is left to future work.
- (c) **Liquid transparency:** CapCam assumes that the liquid is transparent and the pattern at the bottom is visible from the surface. Many liquids are transparent and hence our model directly applies to them. Even

when a liquid is not sufficiently transparent, it can be diluted with water, and the results can be mapped back to undiluted liquid based on dilution level.

- (d) **Phone and camera requirements:** capillary waves travel quickly. Hence, when taking images of the waves, it is important to choose a fast shutter speed. New phone models have an API for configuring the camera shutter speed and exposure parameters, and hence our choice of evaluating CapCam on an iPhone X. In our experiments, we set the shutter speed so that the exposure time is 1/800 s. Different phone models may come with different cameras and flashlights, and hence require a different choice for the shutter speed. Note that the automatic exposure configuration will not work because the phone tends to choose a longer exposure time, which causes the wave pattern to be fuzzy. One also needs to measure the resolution of the camera and the vibration frequency of the vibro-motor. These measurements just need to be done once for each phone model.

9. CONCLUSION

In this article, we introduce CapCam, the first mobile application that can measure liquid surface tension. It is based on the relationship between surface tension and capillary waves, and it is convenient and accurate. Our evaluation shows that CapCam has an absolute surface tension error of only 0.75 mN/m, and based on measured surface tension, it can successfully measure alcohol concentration and detect water contamination. Our experiments also show it is capable of accurately tracking the protein level in urine, a key physiological index used in diabetes and kidney disease management. We believe this work can serve as a useful tool and enable new meaningful applications and interactions.

Acknowledgments

We would like to acknowledge Gareth McKinley for providing access to the digital tensiometer used for ground truth measurements of surface tension. We thank Deepak Vasisht, Guo Zhang, and the members of NETMIT for their insightful discussion and comments. We also thank our shepherd and the anonymous reviewers for their valuable feedback.

References

- Behroozi, F., Perkins, A. Direct measurement of the dispersion relation of capillary waves by laser interferometry. *Am. J. Phys.* 74, 11 (2006), 957–961.
- Blandford, R.D., Thorne, K.S. *Applications of Classical Physics*. California Institute of Technology, California, USA, 2008.
- Bormashenko, E., Musin, A. Revealing of water surface pollution with liquid marbles. *Appl. Surf. Sci.* 255, 12 (2009), 6429–6431.
- Chen, H., Muros-Cobos, J.L., Amirfazli, A. Contact angle measurement with a smartphone. *Rev. Sci. Instrum.* 89, 3 (2018), 035117.
- Chen, H., Muros-Cobos, J.L., Holgado-Terriza, J.A., Amirfazli, A. Surface tension measurement with a smartphone using a pendant drop. *Colloids Surf. A Physicochem. Eng. Asp.* 533, (2017), 213–217.
- Cicconet, M., Birodkar, V., Lund, M., Werman, M., Geiger, D. A convolutional approach to reflection symmetry. *Pattern Recognit. Lett.* 95, (2017), 44–50.
- Dhekne, A., Gowda, M., Zhao, Y., Hassanieh, H., Choudhury, R.R. Liquid: A wireless liquid identifier. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (2018), ACM, Munich, Germany,

- 442–454.
- Diskin, C.J., Stokes, T.J., Dansby, L.M., Carter, T.B., Radcliff, L. Surface tension, proteinuria, and the urine bubbles of hippocrates. *Lancet* 355, 9207 (2000), 901–902.
- FDS. Digital tensiometer – dcat11, 2018. <http://www.fdsc.com/wpdev15/?portfolio=dca11>.
- A. Government. Microalbumin level: 2018. <https://meteor.ahw.gov.au/content/index.phtml/itemId/270339/>.
- Goy, N.A., Denis, Z., Lavaud, M., Grolleau, A., Dufour, N., Deblais, A., Delabre, U. Surface tension measurements with a smartphone. *Phys. Teach.* 55, 8 (2017), 498–499.
- Ha, U., Ma, Y., Zhong, Z., Hsu, T.M., Adib, F. Learning food quality and safety from wireless stickers. In *Proceedings of the 17th ACM Workshop on Hot Topics in Networks* (2018), ACM, Redmond, Washington, USA, 106–112.
- Kruss. Force tensiometer – k6, 2018. <https://www.kruss-scientific.com/products/tensiometers/force-tensiometer-k6/>.
- Liang, C.K., Chang, L.W., Chen, H.H. Analysis and compensation of rolling shutter effect. *IEEE Transactions on Image Processing*, 2008.
- Mogensen, C. Microalbuminuria predicts clinical proteinuria and early mortality in maturity-onset diabetes. *N. Engl. J. Med.* 310, 6 (1984), 356–360.
- Pye, V.I., Patrick, R. Ground water contamination in the united states. *Science* 221, 4612 (1983), 713–718.
- Rahman, T., Adams, A.T., Schein, P., Jain, A., Erickson, D., Choudhury, T. Nutrilizer: A mobile system for characterizing liquid food with photoacoustic effect. In *SenSys*. ACM, Stanford, CA, 2016.
- Speight, J.G., et al. *Lange's handbook of chemistry, Volume 1*. McGraw-Hill, New York, 2005.
- Sridhar, M., Reddy, C.R. Surface tension of polluted waters and treated wastewater. *Environ. Pollut. B. Chem. Phys.* 7, 1 (1984), 49–69.
- Vazquez, G., Alvarez, E., Navaza, J.M. Surface tension of alcohol + water from 20 to 50°C. *J. Chem. Eng. Data* 40, 3 (1995), 611–614.
- Wang, J., Xiong, J., Chen, X., Jiang, H., Balan, R.K., Fang, D. Tagscan: Simultaneous target imaging and material identification with commodity rfid devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, Salt Lake City, Utah, USA, 2017.
- Wei, M., Huang, S., Wang, J., Li, H., Yang, H., Wang, S. The study of liquid surface waves with a smartphone camera and an image recognition algorithm. *Eur. J. Phys.* 36, 6 (2015), 065026.
- Yeo, H.S., Flamich, G., Schrempf, P., Harris-Birtill, D., Quigley, A. Radarcat: Radar categorization for input & interaction. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, Tokyo, Japan, 2016.
- Zhu, F., Miao, R., Xu, C., Cao, Z. Measurement of the dispersion relation of capillary waves by laser diffraction. *Am. J. Phys.* 75, 10 (2007), 896–898.

Shichao Yue and Dina Katabi ({scyue, dk}@mit.edu), Computer Science & Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA.



This work is licensed under a <https://creativecommons.org/licenses/by/4.0/>

Technical Perspective

The Real-World Dilemma of Security and Privacy by Design

By Ahmad-Reza Sadeghi

THE ROMAN HISTORIAN Tacitus (55 A.D.–120 A.D.) once said “the desire for safety stands against every great and noble enterprise.”

In the digital era, providing security and privacy is a noble enterprise, and the entanglement between security and safety systems is increasing. The growing digitization of smart devices has already become an integral part of our daily lives, providing access to vast number of mobile services. Indeed, many people are *glued* to their smart devices. Hence, it seems almost natural to use them in the context of critical emergency and disaster alerts from life-threatening weather to pandemic diseases. However, despite all the convenience they offer, smart devices expose us to many security and privacy threats.

The following paper investigates real-world attacks on the current implementation of Wireless Emergency Alerts (WEA), which constitutes different emergency categories like AMBER Alerts in child-abduction cases, or alerts issued by the U.S. president.

The 3rd Generation Partnership Project (3GPP) standardization body, consisting of seven telecommunications standard development organizations, has specified and released a standard to deliver WEA messages over Commercial Mobile Alert Service (CMAS) in LTE networks. According to the authors, 3GPP made a design choice to provide the best possible coverage for legitimate emergency alerts, regardless of the availability of working SIM cards required for setting up a secure channel to a network base station. However, this realization leaves every phone vulnerable to spoof alerts. Consequently, all modem chipsets that fully comply with the 3GPP standard show the same behavior, that is, fake Presidential Alerts (and other types of alerts) are received without authentication.

The paper applies the art of engineering and demonstrates as well as extensively evaluates a real-world base

station spoofing attack (that is, disguising a rogue base station as genuine). Basically, the attacker sets up its own rogue base station in the vicinity of the victim(s).

The rogue base station will most probably have a better signal strength than benign stations to the victims’ devices, leading the victim’s device to try to connect to the rogue station. While the phone has failed or is just failing to connect to a (malicious) fake base station, the CMAS message will still be received by the device because the standardized protocol allows it. The attack was simulated in a sports arena by utilizing 4×1Watt malicious base stations located outside four corners of the stadium with 90% success rate (coverage of 49,300 from 50,000 seats). This sounds cool *and* creepy.

Critics may question the originality of the attack and the adversary’s motivation. In fact, a system with no or poorly designed security can certainly be compromised sooner or later. Moreover, faking a bomb threat may have a similar impact as using four fake base stations close to a stadium. However, security researchers typically conduct rigorous risk analysis, weighing each potential threat and their impact and interdependencies. For instance, nation state adversaries have the motivation and the capacity to probe and disrupt a national alert system to create chaos and panic.


From a more general perspective, security researchers and practitioners are often faced with trade-offs between security, privacy, and safety that depend on various constraints such as regulations, risk priority analysis, or migration of legacy systems. There are challenging interdisciplinary questions from technological to legal and societal aspects to be considered when designing digital systems and public safety systems, which shows the multifaceted nature and importance of cybersecurity.

Certainly, it is possible to design an alert system with a reasonable level of

security and high coverage. The authors propose multiple countermeasures: end-to-end digital signature on the message; pre-shared cryptographic keys on the phone for authorized entities to issue alerts (for example, the U.S. President, law enforcement); ignoring all alert-specific messages before the mobile device successfully authenticates the network; leveraging configuration information sent by the base station to determine a fingerprint; or using the device’s received signal strength (RSS) to determine if the connected base station is a feasible distance away.

These solutions have their own pros and cons and can work successfully under further assumptions, such as the existence of a public-key infrastructure, or that only devices shipped with the corresponding cryptographic keys can receive these messages, or they require a specific app running in the background on the device.

In summary, this research demonstrates it is non-trivial to design and develop a public safety system that provides security and/or privacy guarantees with high coverage of legacy systems. We can already witness this problem in the context of digital tracing apps deployed to support manual tracing of COVID-19 infection chains. Hence, it is vital that system specifications and standardizations are accompanied with a thorough threat analysis—because of their long-term impact, the compromises we make today may have fatal consequences tomorrow.

The moral of the story is that security and privacy by design are noble enterprises and are crucial in a world becoming increasingly more dependent on “intelligent” digital technologies where security and safety functions are highly intertwined. 

Ahmad-Reza Sadeghi is a professor of computer science at Technische Universität Darmstadt, in Germany, where he heads the System Security Lab.

Copyright held by author.

Securing the Wireless Emergency Alerts System

By Jihoon Lee, Gyuhong Lee, Jinsung Lee, Youngbin Im, Max Hollingsworth, Eric Wustrow, Dirk Grunwald, and Sangtae Ha

Abstract

Modern cell phones are required to receive and display alerts via the Wireless Emergency Alert (WEA) program, under the mandate of the Warning, Alert, and Response Act of 2006. These alerts include AMBER alerts, severe weather alerts, and (unblockable) Presidential Alerts, intended to inform the public of imminent threats. Recently, a test Presidential Alert was sent to all capable phones in the U.S., prompting concerns about how the underlying WEA protocol could be misused or attacked. In this paper, we investigate the details of this system and develop and demonstrate the first practical spoofing attack on Presidential Alerts, using commercially available hardware and modified open source software. Our attack can be performed using a commercially available software-defined radio, and our modifications to the open source software libraries. We find that with only four malicious portable base stations of a single Watt of transmit power each, almost all of a 50,000-seat stadium can be attacked with a 90% success rate. The real impact of such an attack would, of course, depend on the density of cellphones in range; fake alerts in crowded cities or stadiums could potentially result in cascades of panic. Fixing this problem will require a large collaborative effort between carriers, government stakeholders, and cellphone manufacturers. To seed this effort, we also propose three mitigation solutions to address this threat.

1. INTRODUCTION

The Wireless Emergency Alerts (WEA) program is a government mandated service in commercialized cellular networks in the U.S. WEA was established by the Federal Communications Commission (FCC) in response to the Warning, Alert, and Response Act of 2006 to allow wireless cellular service providers to send geographically targeted emergency alerts to their subscribers. The Federal Emergency Management Agency (FEMA) is responsible for the implementation and administration of a major component of WEA.

This system can send three types of alerts: **Presidential Alerts** issued by the president to all of the United States; **Imminent Threat Alerts** involving serious threats to life and property, often related to severe weather; and **AMBER Alerts** regarding missing or abducted children. Considering the number of cellphone users and the nationwide coverage of cellular networks, WEA over Long-Term Evolution (LTE) was a natural step to enhance public safety *immediately* and *effectively*. In fact, recent rapidly moving fires have caused emergency services to consider using WEA instead of relying on opt-in alerting systems.¹⁶

A handful of widely publicized events has led to public scrutiny over the potential misuse of the alert system. On

January 13, 2018, there was a geographically targeted alert issued in Hawaii. The message, warning of an inbound missile, is shown in Figure 1b. Although caused by human error, the impact on the residents of Hawaii was huge, as it led to panic and disruption throughout the state.²⁰ This event was followed on October 3, 2018, by the first national test of a mandatory Presidential Alert. The alert, captured in Figure 1a, was sent to all capable phones in the U.S.¹⁹

These recent high-profile alerts have prompted us to assess the realizability and impact of an alert spoofing attack. In this paper, we demonstrate how to launch a Presidential Alert-spoofing attack and evaluate its effectiveness with respect to attack coverage and success rate.

To answer this question, we start by looking into the alert delivery method used by WEA. WEA sends alerts via the commercial mobile alert service (CMAS), which is the underlying delivery technology standardized by the 3rd Generation Partnership Project (3GPP). These alerts are delivered via the LTE downlink within broadcast messages, called System Information Block (SIB) messages. A celltower (referred to as eNodeB) broadcasts the SIB to every cell phone (referred to as user equipment or UE) that is tuned to the control channels of that eNodeB. A UE obtains necessary access information, such as the network identifier and access restrictions, from SIB messages, and uses it for the eNodeB selection procedure. Among the 26 different types of SIB messages, SIB12 contains the CMAS notification, which delivers the aforementioned alert messages to the UEs.

The eNodeB broadcasts SIB messages to the UE, independently from the mutual authentication procedure that

Figure 1. Snapshots of real WEA messages received by cellphones: (a) the first national test of the Presidential Alert performed on October 3, 2018 in the U.S., and (b) a false alert sent in Hawaii on January 13, 2018.



(a) Presidential alert

(b) Imminent threat alert

The original version of this paper is entitled “This is Your President Speaking: Spoofing Alerts in 4G LTE Networks” and was published in *Proceedings of the 17th ACM International Conference on Mobile Systems, Applications and Services*, 2019.

eventually occurs between them. Thus, all SIBs, such as CMAS, are intrinsically *vulnerable* to spoofing from a malicious eNodeB. More importantly, even if the UE has completed its authentication and securely communicates with a trusted eNodeB, the UE is still exposed to the security threat caused by the broadcasts from other, possibly malicious, eNodeBs. This is because the UE periodically gathers SIB information from neighboring eNodeBs for potential eNodeB (re)selection and handover.

We found via both experiment and simulation that a 90% success rate can be reached in 4435 m² of a 16,859 m² building using a single malicious eNodeB of 0.1 Watt power, whereas in an outdoor stadium, 49,300 seats among the total 50,000 are hit with an attack, which itself has a 90% success rate using four malicious eNodeBs of 1 Watt power.

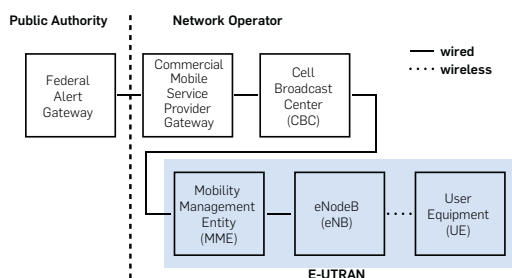
In summary, we make following major contributions:

- We identify security vulnerabilities of the WEA system and explain the detailed underlying mechanism stipulated by the LTE standard. We find that the CMAS spoofing attack is easy to perform but is challenging to defend in practice.
- We present our threat analysis on the CMAS spoofing attack and implement an effective attack system using commercial off-the-shelf (COTS) software-defined radio (SDR) hardware and open-source LTE software.
- We evaluate our attack system using both SDR-based hardware prototype and measurement-based simulation. As one of the striking results, we demonstrate that four SDR-based malicious eNodeBs at 1 Watt of power can propagate their signal to 49,300 of the whole 50,000-seat football stadium. Of the 49,300 seats affected, 90% will receive the CMAS message.
- We present possible solutions to prevent such a spoofing attack with a thorough analysis and feasibility test, which can open the door toward collaborative efforts between cellular operators, government stakeholders, and phone manufacturers.

1.1. Responsible disclosure

In January 2019, before public release, we disclosed the discoveries and technical details of this alert spoofing attack to various pertinent parties. These parties include the government and standardization organizations FEMA, FCC, DHS, NIST, 3GPP, and GSMA; the cellular network service

Figure 2. LTE CMAS network architecture.



providers AT&T, Verizon, T-Mobile, Sprint, and U.S. Cellular; and the manufacturers Apple, Google, and Samsung.

2. SECURITY THREATS

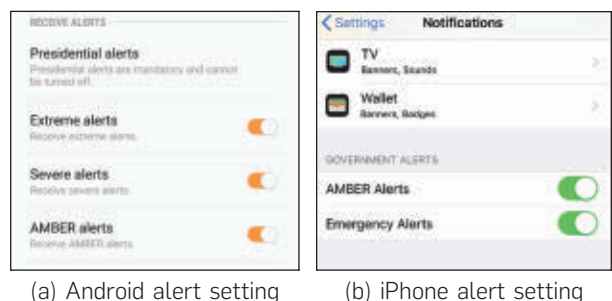
The 3GPP standardization body began a project in 2006 to define the requirements of CMAS to deliver WEA messages in the LTE network, and the LTE CMAS network architecture is illustrated in Figure 2. The resulting technical specification, initially released in 2009, describes the general criteria for the delivery of alerts, message formats, and functionality of CMAS-capable UEs.² During an emergency, authorized public safety officials send alert messages to Federal Alert Gateways. The participating mobile service providers then broadcast the alert to the UEs, who will automatically receive the alert if they are located in or travel to the targeted geographic area. The cell broadcast center (CBC) is part of the service provider’s core network and is connected to the Mobility Management Entity (MME), which maintains the location information of the UEs attached to the network.³ The eNodeB is the final step in communicating the alert to the UEs over the air.

UEs may choose to turn off the notification of imminent threat alerts and AMBER alerts among the three types of emergency alerts (i.e., presidential alerts, imminent threat alerts, and AMBER alerts). However, the 3GPP mandated that the reception of Presidential Alerts is obligatory. Thus, cell phones have no option to disable Presidential Alerts, as seen in Figure 3. Because it cannot be disabled, this paper focuses on spoofing Presidential Alerts with the injection of a fake CMAS message over the air from a rogue eNodeB.

2.1. Identifying the vulnerability

An eNodeB broadcasts LTE system information through the Master Information Block (MIB) and SIB. Specifically, when a LTE searches for an eNodeB, it searches for the eNodeB’s physical cell identifier (PCI) within a dedicated synchronization channel specified by the LTE standard.⁵ After finding the PCI, the LTE unscrambles the MIB, which contains essential information such as the system bandwidth, system frame number (SFN), and the antenna configuration, to decode the SIB Type 1 message (SIB1). There are several SIB messages but only SIB1 has a fixed periodicity of 80

Figure 3. Government alert settings in mobile phones: (a) Android and (b) Apple’s iPhones. Although AMBER and emergency alerts can be manually disabled, users cannot disable or block Presidential Alerts from being received or displayed.



msec. Other SIB messages are dynamically scheduled by the eNodeB, and the scheduling information for other SIBs is encoded in the periodic SIB1.

3GPP specifies that the broadcast of CMAS messages is over the air through SIB12.⁶ However, unlike point-to-point messages in LTE, broadcasts of SIB messages are not protected by mutual cryptographic authentication or confidentiality, because the SIB contains essential information the UEs use to access the network before any session keys have been established. Once a CMAS message has been received, there is no verification method for the message content. If an attacker can imitate eNodeB behavior closely enough to broadcast false CMAS messages, the UE will display them.

A UE's vulnerability to a fake CMAS alert depends on whether it is in an *active* or *idle* state, illustrated in Figure 4. To affect the most UEs, an attacker must consider different approaches for each state. Here we discuss idle UEs and active UEs separately:

Idle mode UEs. Reference Signal Received Power (RSRP) is the power of an eNodeB-specific reference signal recognized by the UE, typically used to make an eNodeB selection and handover decision. Usually, whenever a UE in idle mode performs eNodeB selection (or reselection), it will associate with the eNodeB having the highest RSRP. If the RSRP of a malicious eNodeB is the strongest, the UE decodes the SIBs transmitted by the malicious eNodeB. The attacker does not need to have any user information (such as security keys), which would be stored in the network operator's database. Without having such user information, the UE will eventually reject the authentication process with the malicious eNodeB. However, it can receive a CMAS message transmitted by the malicious eNodeB during this process.

Active mode UEs. When a UE is in active mode, it securely communicates with the serving eNodeB. If it finds another eNodeB with a higher power level than the existing serving eNodeB, a handover procedure can be triggered. The serving eNodeB then makes a handover decision based on the received measurement report. However, if the serving MME

does not identify the target eNodeB, the handover will eventually fail. Therefore, even if caused by a malicious eNodeB, the handover procedure does not make a UE vulnerable to the CMAS spoofing attack. As a consequence, the attacker first needs to disconnect the UE from its serving eNodeB. After the UE is released from the serving eNodeB, it will immediately try to attach to the strongest eNodeB. After that, it can be attacked in the same way as idle mode UEs described in the section above. One way to disconnect the active UE from its serving eNodeB is to incur Radio Link Failures (RLFs) by jamming LTE signals.¹⁵ Simply, without any special jamming technique, a malicious eNodeB can jam the communication between a UE and its serving eNodeB by merely transmitting at a much higher power than the serving eNodeB.

2.2. CMAS reception and trustworthiness

We have identified three possible cases that determine whether the CMAS is received and is trustworthy in Table 1. Each case depends on where the UE is currently in the idle/active life cycle, illustrated in Figure 4.

Simply put, if a UE is not listening to frequency channels on which the eNodeB is transmitting the CMAS message, the CMAS message will not be received by the UE. This is illustrated as the blue portion in Figure 4. It may seem obvious, but a necessary condition for the UE to receive a CMAS message is that it needs to be tuned to the synchronization channels of the eNodeB that is transmitting the CMAS message.

Secure CMAS. In the green area of Figure 4, the UE attaches to an eNodeB and is safely in the active state. To do this, the UE must be equipped with a valid Service Identity Module (SIM) card that is registered to the operator's network. Case 1 is the general scenario for phones receiving standard service from their provider. Because mutual authentication between the UE and the network has been successfully made, the UE can trust that the eNodeB is not malicious. The CMAS reception is successful as we would expect, and we know that this CMAS message is trustworthy.

Unsecured CMAS. In the red area of Figure 4, the UE is failing or has already failed to attach when the eNodeB transmits the CMAS message. The UE will still receive the CMAS message; this is the crux of the vulnerability. To demonstrate this, we deleted the SIM information from the Evolved Packet Core (EPC) so that the user authentication would be unsuccessful. The UE is now in the unsecured range between the idle and active states due to the authentication failure. Even though the UE fails to reach the active state, we observe that the CMAS message is still successfully received. This is because once the UE completes decoding the CMAS message in SIB12, it delivers the contents to the application layer to be shown to the user. Surprisingly, this is possible even after

Figure 4. The Idle/Active life cycle of a UE. The state of the UE continues counterclockwise around the chart. CMAS spoofing is possible although the UE performs an eNodeB search, prior to successful authentication with a trusted eNodeB.

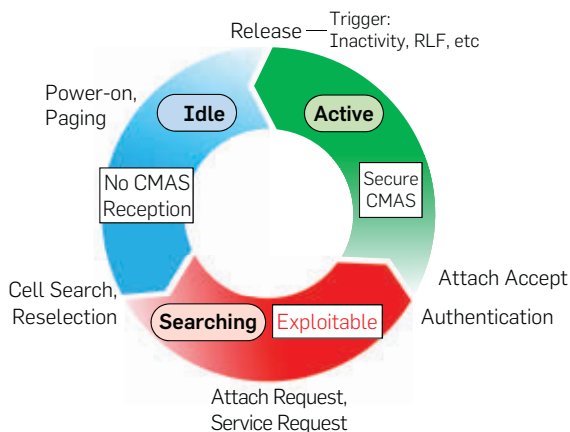


Table 1. Cases for CMAS reception and trustworthiness.

Case	SIM equipped	Auth. success	CMAS reception	Trustworthy
1	Yes	Yes	Yes	Yes
2	Yes	No	Yes	No
3	No	No	Yes	No

the authentication process has finally failed. Case 2 can lead the potential threat that *any malicious eNodeB can deliver fake CMAS messages although the UE is in between the eNodeB search and authentication procedures*. Finally, in Case 3, the UE roams to an eNodeB, which sends a CMAS message. To demonstrate this, we removed the SIM card from the UE. No authentication is possible, but the UE can make emergency calls such as 911. Even in this situation, we verified that the UE still receives the CMAS message, which is potentially malicious.

As shown in Cases 2 and 3, CMAS spoofing can be done although the UE performs an eNodeB search before successful authentication with a trusted eNodeB. These results are verified using 1 × JL620 COTS LTE small cell (no modification), 1 × open-source NextEPC (modified with the CBC),¹⁷ and nine different commercial LTE phones (Apple iPhone 8, X, and XS; Google Pixel 1; Huawei Nexus 6P; Motorola G5 Plus and G6; Samsung Galaxy S7 Edge and S8). Considering that the majority of UEs in cellular networks are in the idle state¹⁰ and UEs often transition from the active to idle state due to an inactivity timer (around 10 s¹³), *almost all UEs are susceptible to this attack*.

3. PROOF-OF-CONCEPT ATTACKS

In this section, we present the details of our *Presidential Alert Spoofer* system and describe how it works. Our system can be built with either an SDR device or a COTS eNodeB, and the list of hardware and software systems we used is summarized in Table 2.

Attack preparation. Our Presidential Alert Spoofer must first identify the existing eNodeBs in a given licensed frequency band. Each eNodeB can be uniquely identified at a given geographical position by the pair of “E-UTRA Absolute Radio Frequency Channel Number (EARFCN)” and “Physical Cell ID (PCI).” For each EARFCN, our Spoofer finds the eNodeB, and associated PCI, of which the RSRP is the strongest. Once the existing eNodeBs are listed, the Public Land Mobile Network (PLMN) information of each eNodeB is collected. Every LTE network has its PLMN, a three-digit country code, and two or three digits to identify the provider. The PLMN is periodically broadcast by the eNodeB in the SIB1 message, making it possible to collect all of the observable PLMNs within the receiving range passively. To launch an attack, our Presidential Alert Spoofer uses the same PLMN as an existing eNodeB such that the UEs will select our Spoofer during an eNodeB search.

Table 2. HW and SW systems used for implementation.

System	Hardware	Software
Attack preparation	BladeRF 2.0 (\$500) USRP B210 (\$1300) Laptop (< \$1000)	OWL ⁹ (modified)
SDR-based Spoofer	BladeRF 2.0 (\$500) USRP B210 (\$1300) Laptop (< \$1000)	srsLTE ¹² (modified)
COTS eNodeB-based Spoofer	JL620 (FDD) JLT621 (TDD) Laptop (< \$1000)	NextEPC ¹⁷ (modified)

Attack execution with an SDR device. We implemented the Spoofer using a USRP B210 and BladeRF to attack Frequency Division Duplex (FDD) systems. With an SDR, we can change the transmission frequency easily to target every cellular band. We added SIB12 support to the open-source eNodeB software¹² and could transmit a CMAS message every 160 msec.

Attack execution with a COTS eNodeB. We use a COTS eNodeB (Juni JLT-621) to target Time Division Duplex (TDD) systems. Our modification of NextEPC provides an interface to inject a user-defined Presidential Alert that broadcasts each second. With this configuration, a victim UE may receive the SIB12 every second from the COTS eNodeB. Any commercial LTE FDD/TDD eNodeB hardware can perform this attack, which may play a key role if an attacker wants to control multiple malicious eNodeBs in a coordinated manner.

Attack verification. In our lab environment, we verified that the fake Presidential Alert sent by our SDR-based

Figure 5. The Presidential Alert Spoofer scans for an eNodeB, gathers operator information, and sends a fake Presidential Alert to both idle and active UEs. The UEs may be FDD or TDD. This setup consists of one SDR device, one COTS LTE eNodeB, and two laptops.

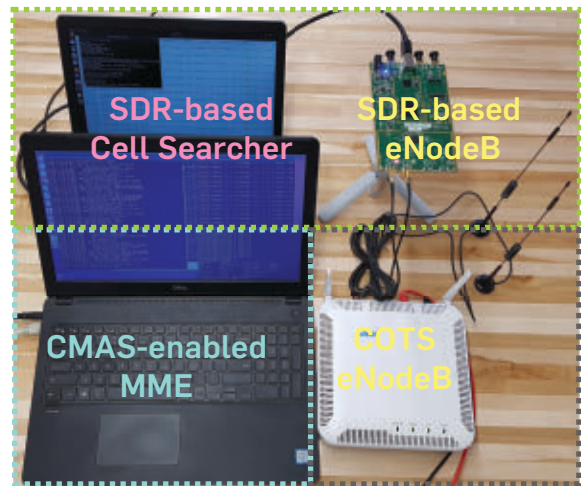
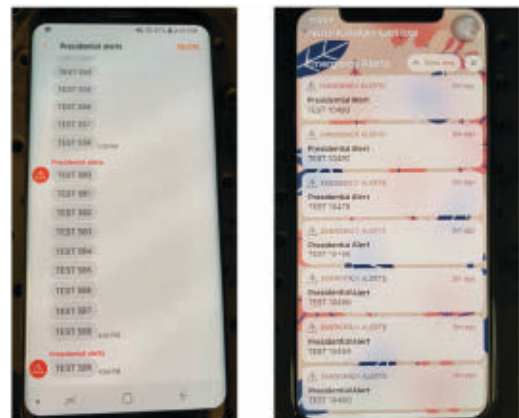


Figure 6. Receiving multiple fake Presidential Alerts using a Samsung Galaxy S8(left) and an Apple iPhone X(right).



Spoofers were successfully shown in the FDD phones of AT&T, T-Mobile, and Verizon. With a TDD Sprint phone, we verified that our COTS eNodeB-based Spoofer also works successfully. All the experiments are carried out with proper RF shielding.

Affected devices and implications. From discussions of the SIB12 vulnerability in §2.1, it became clear that the lack of authentication was a design choice by 3GPP, rather than an oversight. This design provides the best possible coverage for legitimate emergency alerts, but the trade-off leaves every phone vulnerable to spoofed alerts. Consequently, all modem chipsets that fully comply with the 3GPP standards show the same behavior: the fake Presidential Alert is received without authentication. Once the LTE modem of the UE receives the fake alert, the operating system will display the alert to the user. Because our attack verification tests included many Android and iOS phones, we conclude that most (presumably all) LTE phones will be affected by the attack, regardless of the phone's vendor or model. Moreover, much of the LTE public warning system is inherited from 2G/3G and continues in 5G; a similar attack is also possible in 5G.

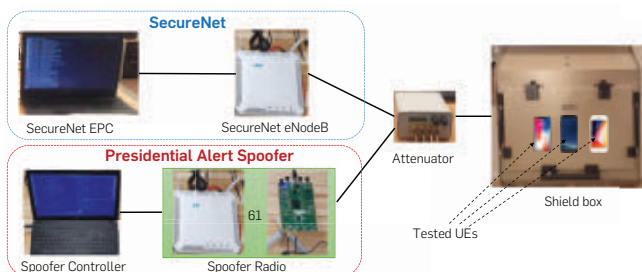
4. EVALUATION

Figure 7 illustrates our experimental testbed setup, which consists of an EPC and eNodeB for a conventional LTE system, a malicious eNodeB for spoofing, and cell phones for victim UEs. A signal attenuator receives the broadcast signals from two sources and delivers the combined signal to a LTE in a shielded box. We built an LTE test network with an EPC and eNodeB, named SecureNet, which assumes the role of the user's original network. On the other hand, the malicious eNodeB, part of the Presidential Alert Spoofer, is installed solely without any LTE core support. By using the signal attenuator, the signal power received at the LTE can be precisely controlled for various practical scenarios.

4.1. Success rate

Let α be the RSRP difference between the SecureNet eNodeB and Presidential Alert Spoofer for an idle UE (i.e., $\alpha = RSRP_{SecureNet} - RSRP_{Spoofer}$) and β be the RSRP difference for an active UE. Then we evaluate the Presidential Alert Spoofer's success rate as a function of α (or β). We first attach the UE to SecureNet. For the idle UE case, we wait for

Figure 7. The testbed setup for evaluating the attack success rate. The transmission power levels of the SecureNet eNodeB and the Presidential Alert Spoofer can be controlled independently.



the UE to enter the idle mode due to inactivity. The Spoofer broadcasts each new Presidential Alert message, so we can determine whether each Presidential Alert is successfully received and at what power configuration of α or β . We conducted 20 experimental trials for each value of α (or β) ranging from 0 to -25 dB.

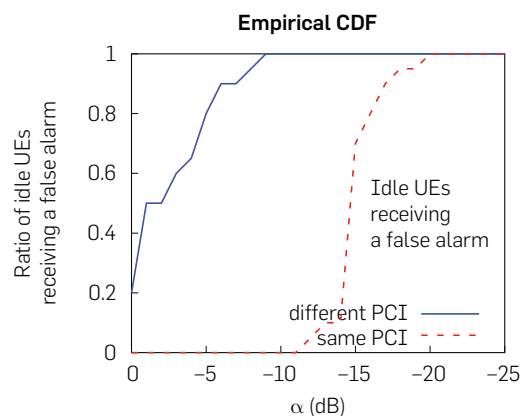
The Spoofer may elect to use a different PCI than that of the serving eNodeB, appearing to be a new eNodeB. Or, the Spoofer may use the same PCI, looking to be the existing eNodeB and interfering with the existing eNodeB's PHY-layer control channel information.²² This decision has different impacts on the performance of the spoofing attack, depending on the UE state (*idle* or *active*).

Figure 8 shows the empirical cumulative distribution function (CDF) of successful receptions of fake alerts as a function of α for idle UEs. When the Spoofer uses a different PCI and the received signal strength from the Spoofer is higher than that from SecureNet ($\alpha < 0$), the idle UE will consider the Spoofer as a new serving eNodeB. Our experimental results verify this expectation; 50% of idle UEs can receive a fake message even at $\alpha = -1$, and more than 90% of idle UEs can receive a fake message when $\alpha \leq -6$.

However, if the same PCI is used, the attack performance is significantly degraded. Because the PCI is used to generate cell-specific reference signals,⁵ using the same PCI value will cause channel estimation errors at the UE due to collisions from the two transmitters. This, in turn, leads to more decoding errors when receiving the SIBs. As a result, using the same PCI requires much higher attack power as no UE is affected when α is greater than -12 dB. With $\alpha \leq -17$, 90% of idle UEs can still be attacked.

Figure 9 shows the CDF of successful fake message receptions as a function of β (i.e., forcing disconnect) for active UEs. When the Spoofer uses a different PCI, and the received signal strength from the Spoofer is higher than that from the SecureNet eNodeB, the active UE will start to consider the Spoofer as a target eNodeB for a handover, as described in §2.2. Because SecureNet does not identify the Spoofer, a handover cannot be performed. Instead, we observed an RLF would occur when $\beta \leq -10$, which eventually leads to the

Figure 8. The CDF as a function of α for only idle UEs. Because eNodeB reselection happens when idle UEs wake up, the spoofing attack performs better when using a different PCI.



reception of a fake alert. About 90% of active UEs can receive a fake message when $\beta \leq -20$, assuming that a different PCI value is used for the Spoofer. Unlike the idle UE case, using the same PCI value results in higher decoding errors (and more RLFs) at a receiver. Thus, it shows better attack performance; 90% of receptions are successful with $\beta \leq -13$.

4.2. Practical scenarios: indoor and outdoor

As we do not use the Spoofer outside of a shield box, we cannot directly measure its effect on a large number of people. To evaluate the attack coverage according to its success rate, we use actual RSRP measurements in indoor and outdoor environments.

Indoor attack. We placed our malicious eNodeB inside a campus building and measured the RSRP of a dummy LTE signal (containing no CMAS message) in the EBS band with 0.1 Watt transmit power. We also measured the RSRP of a nearby AT&T eNodeB, as shown in Figure 10a. The RSRP does not attenuate consistently due to various obstacles, but generally, the RSRP tends to decrease as the distance from

Figure 9. The CDF as a function of β for only active UEs. Using the same PCI leads to more decoding errors observed by the UE. This results in a slightly more effective attack.

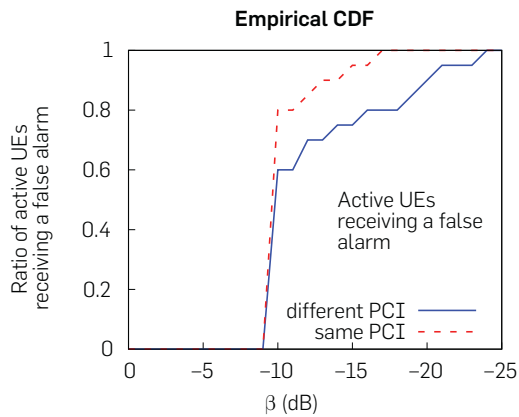
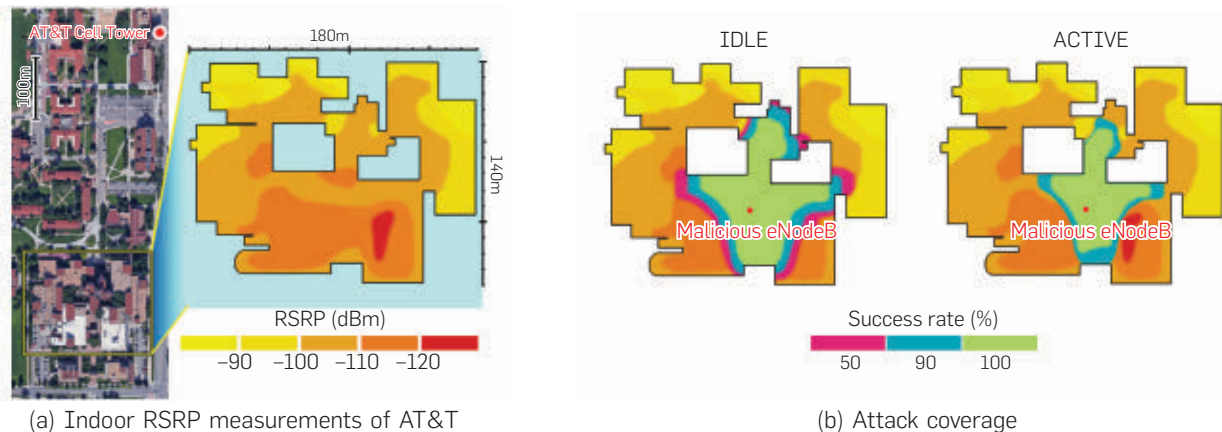


Figure 10. The indoor attack simulation: (a) the satellite image of the Engineering Center at the University of Colorado Boulder shows the nearest AT&T eNodeB. The graph shows the indoor RSRP distribution of that eNodeB. (b) The attack coverage for idle and active UEs are shown when a 1×0.1 Watt malicious eNodeB is used.



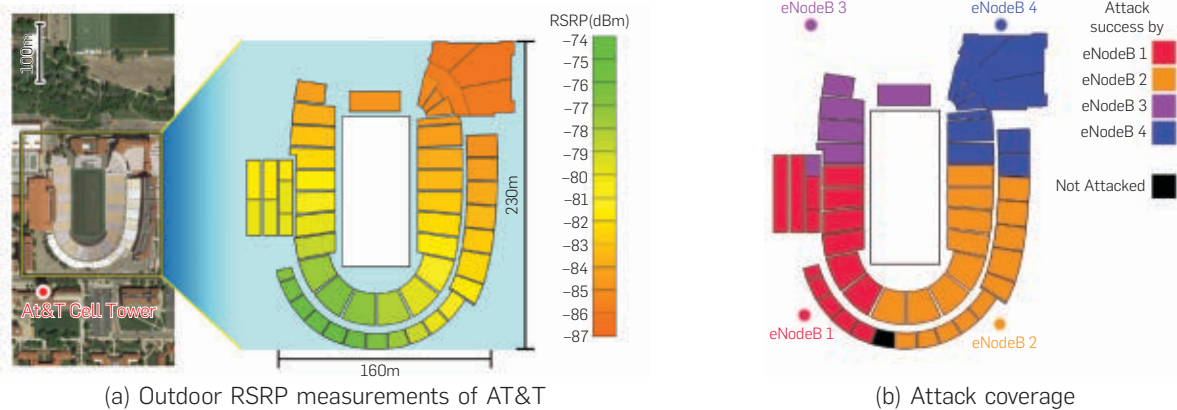
the AT&T eNodeB increases. We compared the two RSRPs throughout the building and indicated the attack coverage using measurements obtained from §4.1, as depicted in Figure 10b. As a result, in a building with a total area of about 16,859 m², for idle UEs, the coverage for a 90% success rate was about 4435 m², whereas for active UEs, the coverage for a 90% success rate was about 2955 m².

Outdoor attack. Without access to outdoor LTE equipment, we simulate the RSRPs of the spoofing eNodeB and the AT&T eNodeB with the NS-3 v3.29 network simulator.¹⁸ For the scenario, we assume a football game where a large number of people are gathered in a restricted region. A group of attackers sends fake alerts to the spectators inside the football stadium. We measured the RSRP of an actual AT&T eNodeB around the perimeter of our campus' football stadium, as shown in Figure 11. We used the simulator to estimate the RSRPs at the centers of each section in the stadium (Figure 11a). We simulated the spoofer in four corners around the stadium, near but still outside of the ticketed area. Figure 11b shows which malicious eNodeB with a 1 Watt transmit power attacked each section. We observe that all sections, except one, are attacked by the malicious eNodeBs. This means that 49,300 among the total 50,000 seats will be hit with the attack, which itself has a 90% success rate, given that all UEs are in the idle state.

5. MITIGATION SOLUTIONS

Defending against CMAS spoofing attacks requires careful consideration of several challenges. First, updates to the CMAS architecture could require expensive changes by cell phone manufacturers, operating system developers, government bodies, and cellular carriers. Coordinating such an effort would be difficult due to the fragmented nature of the network. Furthermore, updates must still support outdated devices, both on the user (UE) and infrastructure (eNodeB) side, as it could take years to replace old equipment. Also, any comprehensive defense must consider the trade-off between security and availability: if users cannot receive valid alerts due to sophisticated protections, it may be more

Figure 11. The outdoor attack simulation: (a) the satellite image of Folsom Field at the University of Colorado Boulder shows the location of the AT&T eNodeB. The stadium graph represents the RSRP distribution of the eNodeB measured at the center of each section, (b) When 4 × 1 Watt malicious eNodeBs are located outside the four corners of the stadium, the simulated attack coverage hits all but one section. This means that 49,300 among the total 50,000 seats are hit with the attack, which itself has a 90% success rate.



hazardous than the case if we continued to use the existing (but vulnerable) system.

With these challenges in mind, we propose three mitigation solutions: first, a client-side software solution ignoring unsecured CMAS alerts; second, a network-aware solution attempting to detect false alerts by modeling characteristics of legitimate eNodeBs; and third, adding digital signatures to alerts.

5.1. Client-driven approach

A client-driven approach should provide an ability for a UE to decide whether a received CMAS message is trustworthy. It requires the information from LTE’s control plane, which is responsible for essential operations such as network attaches, security control, authentication, setting up of bearers, and mobility management. To mitigate the CMAS spoofing attack, we utilize Radio Resource Control (RRC) and Non-Access Stratum (NAS) layer information from the LTE control plane. We can check whether the UE has a valid connection or not from the RRC control information and the UE’s state transition with MME from the NAS control information.

Monitoring RRC and NAS on a UE is currently tricky because LTE control plane protocols are handled by the LTE baseband chipset and firmware so that accessing such information through the existing Operating System (e.g., Android, iOS) is not fully supported. In our implementation, we installed a cellular debugging tool on Android to retrieve the state information of RRC and NAS.¹⁴

Figure 12 shows the RRC and NAS state transition in a standard scenario where the UE receives a *Secure CMAS* message from a legitimate eNodeB. When it receives an *Unsecure CMAS* message, we will see a different state transition. For instance, when a UE is in active mode, the attack starts with a sudden radio link failure, as we explained in §2. It incurs the RRC state change from “CONNECTED” to “IDLE,” and the state goes back to “CONNECTED” when a CMAS is received. After that, the NAS state will soon change into “EMM-REGISTERED.NO-CELL-AVAILABLE.”

Figure 12. UE state transition for Secure CMAS reception.

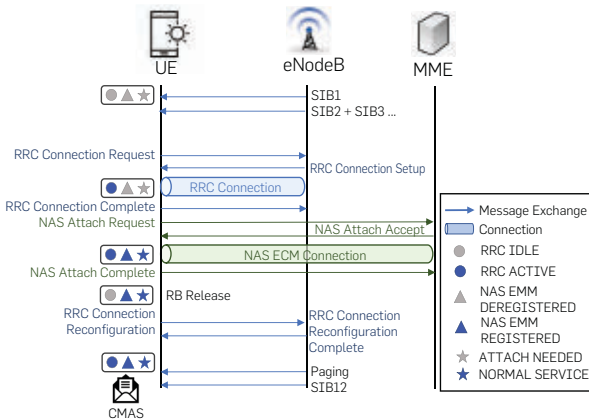
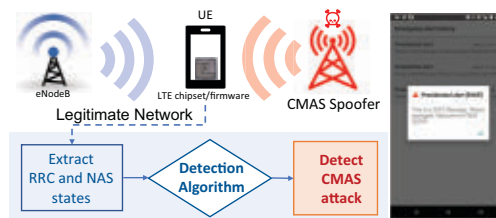


Figure 13. The client-driven approach evaluates the security of the broadcast radio channel by monitoring UE’s RRC and NAS state transitions when a CMAS message is received. a fake CMAS message with a warning, as shown in Figure 13.



As a result, we propose a spoofing detection algorithm as a client-driven approach. First, it needs to have the ability to access a short history of RRC and NAS state transition. Then, whenever a CMAS message is received, it should be checked by our algorithm before displaying it to a user. The algorithm finds any suspicious activity by evaluating RRC and NAS state transition, assuming that unsecured connections may deliver fake broadcast messages. Finally, it shows:

5.2. Network-aware approach

A network-aware approach can leverage the received signal strength (RSS) at the UE to determine if the eNodeB from which the UE received the CMAS message is within a feasible distance. Using a widely used path-loss model,¹¹ we can estimate the distance to the eNodeB using the RSS value. Then compare this with the location provided by an Internet database⁹ to determine whether the alert could have come from a trusted eNodeB.

The performance of this technique could be greatly improved by applying a machine learning (ML) as shown in Figure 14. In our design, we train legitimate cells using basic cell information, neighbor relations, and signal quality measurements associated with the location. Such information may be collected and shared by network operators or crowdsourcing.²¹ In our prototype, a UE retrieves an ML model associated with its serving and surrounding cells of its location to classify the validity of the attached eNodeB upon reception of a CMAS message.

5.3. Digital signature approach

We also consider digitally signing SIB12 messages to prevent spoofed messages, as discussed by 3GPP.¹ Although it is conceptually simple, adding signatures is difficult because operators and devices must agree on the key or keys that will be used to sign and validate messages.

For key management, we leverage suggestions from 3GPP discussions,¹ which suggest using 1) the Non-Access Stratum (NAS) to send authenticated messages to the device, or 2) Over-The-Air (OTA) UE SIM card provisioning. Because NAS provides message integrity between the eNodeB and UE (mediated by pre-shared keys in the UE SIM card), messages received in this way cannot be spoofed by a (physically) nearby adversary. However, sending alerts over this channel would limit their reception *only* to devices that had established a NAS session. Instead, we recommend using this authenticated channel to send and update a public key that a device should trust. This key should correspond to the private key held by a network operator's Cell Broadcast Center (CBC), which is authorized to broadcast such alerts. Alternatively, the public key distribution can be done using OTA management,⁴

Figure 14. The network provides a machine learning (ML)-based model which characterizes legitimate eNodeBs, and therefore UE can determine whether the alert could have come from a trusted eNodeB or not.

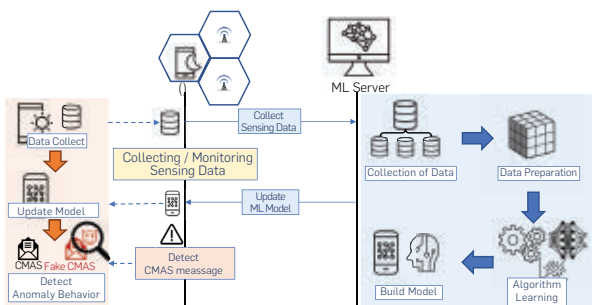
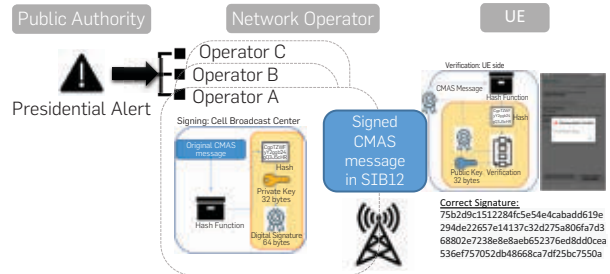


Figure 15. Secure CMAS delivery is guaranteed by adding a signature to alerts. As of May 2019, the FCC mandated to support alert messages up to 360 characters; adding a 64-byte digital signature now becomes applicable for the existing and future wireless emergency alert systems.



which is a well-established technique for updating data on the Universal Integrated Circuit Card (UICC).

To verify this scheme's feasibility, we first stored a public key in a SIM card, assuming that a network operator will provision it. Then we implement the ed25519 digital signature for the Presidential Alert⁷ to sign a 4-byte time stamp along with the CMAS alert message (68 bytes overhead in total). Once a signed message is received, the alert message can be displayed after verifying its signature, as shown in Figure 15. As a result, the UE is not affected by the spoofing attack because it only accepts signed messages.

6. CONCLUSION

In this paper, we have identified the WEA security vulnerabilities over commercial LTE networks and found that a spoofing attack with fake alerts can be made very easily. Specifically, we presented our threat analysis on the spoofing attack and implemented an effective attack system using COTS SDR hardware and open-source LTE software. Our extensive experimentation confirmed that the CMAS spoofing attack could succeed in all tested smartphones in the top four cellular carriers in the U.S. Further, we have proposed potential defenses, from which we believe that completely fixing this problem will require a large collaborative effort between carriers, government stakeholders, and cellphone manufacturers. □

References

- 3GPP TR 33.969. Technical Specification Group Services and System Aspects; Study on security aspects of public warning system (PWS) (Release 15), 2018. <http://www.3gpp.org/DynaReport/33969.htm>.
- 3GPP TS 23.041. Technical Specification Group Core Network and Terminals; Technical realization of Cell Broadcast Service (CBS) (Release 15), 2018. <http://www.3gpp.org/dynareport/23041.htm>.
- 3GPP TS 29.168. Technical Specification Group Core Network and Terminals; Cell Broadcast Centre interfaces with the evolved packet core (Release 15), 2018. <http://www.3gpp.org/dynareport/29168.htm>.
- 3GPP TS 31.115. Technical Specification Group Core Network and Terminals; Secured packet structure for (Universal) subscriber identity module (U)SIM toolkit applications (Release 15), 2019. <http://www.3gpp.org/dynareport/31115.htm>.
- 3GPPx TS 36.211. Technical Specification Group Radio Access Network; Physical channels and modulation (Release 15), 2018. <http://www.3gpp.org/dynareport/36211.htm>.
- 3GPP TS 36.331. Technical Specification Group Radio Access Network; Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC) (Release 15), 2018. <http://www.3gpp.org/dynareport/36331.htm>.
- Bernstein, D.J., Duif, N., Lange, T., Schwabe, P., Yang, B.-Y. High-speed high-security signatures. *J. Cryptographic Eng* 2, 2 (2012), 77–89.
- Bui, N., Widmer, J. OWL: a reliable online watcher for LTE control channel measurements. In *ACM All*

Things Cellular (MobiCom Workshop) (November 2016).


9. CellMapper. Cellular coverage and tower map, 2018. <https://www.cellmapper.net>.
10. Chen, X., Jindal, A., Ding, N., Hu, Y.C., Gupta, M., Vannithamby, R. Smartphone background activities in the wild: origin, energy drain, and optimization. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (2015), MobiCom'15, Paris, France.
11. Goldsmith, A. *Wireless Communications*. Cambridge University Press, Cambridge, England, August 2005.
12. Gomez-Migueluez, I., Garcia-Saavedra, A., Sutton, P.D., Serrano, P., Cano, C., Leith, D.J. srsLTE: an open-source platform for LTE evolution and experimentation. In *ACM WiTECH (MobiCom, Workshop)* (October 2016).
13. Huang, J., Qian, F., Gerber, A., Mao, Z.M., Sen, S., Spatscheck, O. A close examination of performance and power characteristics of 4G LTE networks. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services* (2012), MobiSys'12, Low Wood Bay, Lake District, UK.
14. Li, Y., Peng, C., Yuan, Z., Li, J., Deng, H., Wang, T. Mobileinsight: extracting and analyzing cellular network information on smartphones. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking* (2016), MobiCom'16, New York City, New York, USA.
15. Lichtman, M., Jover, R.P., Labib, M., Rao, R., Marojevic, V., Reed, J.H. LTE/LTE-A jamming, spoofing, and sniffing: threat assessment and mitigation. *IEEE Commun. Mag.* 54, 4 (April 2016), 54–61.
16. National Public Radio. Officials assess response to camp fire in northern california, 2018. <https://goo.gl/IF12Vo>.
17. NextEPC Inc. Open source implementation of LTE EPC, 2019. <https://www.nextepc.com/>.
18. Nsnam. NS-3: a discrete-event network simulator for internet systems, 2018. <https://www.nsnam.org>.
19. The Washington Post. Cellphone users nationwide just received a 'Presidential Alert.' Here's what to know, 2018. <https://goo.gl/KRfDjf>.
20. Wikipedia. Hawaii false missile alert, 2018. <https://goo.gl/oD9ofx>.
21. Yang, D., Xue, G., Fang, X., Tang, J. Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In *The 18th Annual International Conference on Mobile Computing and Networking* (August 2012), MobiCom'12, Istanbul, Turkey.
22. Yang, H., Huang, A., Gao, R., Chang, T., Xie, L. Interference self-coordination: a proposal to enhance reliability of system-level information in OFDM-based mobile networks via PCI planning. *IEEE Trans. Wirel. Commun.* 13, 4 (April 2014), 1874–1887.

Jihoon Lee, Jinsung Lee, Max Hollingsworth, Eric Wustrow, Dirk Grunwald, and Sangtae Ha (jihoon.lee-1, jinsung.lee, max.hollingsworth, ewust}@colorado.edu, dirk.grunwald, sangtae.ha), University of Colorado Boulder, Colorado, USA.

Gyuhong Lee ((caixy))@mnd.go.kr, Korea Army Academy, Yeongcheon, South Korea.

Youngbin Im ((ybim))@unist.ac.kr, UNIST, South Korea.

Jinsung Lee is the corresponding author.

This work is licensed under a  <https://creativecommons.org/licenses/by/4.0/>



Association for Computing Machinery

Career & Job Center

The #1 Career Destination to Find Computing Jobs.




Connecting you with top industry employers.

Your next job is right at your fingertips. Get started today!

The new ACM Career & Job Center offers job seekers a host of career-enhancing benefits, including:

-  Access to new and exclusive career resources, articles, job searching tips and tools.
-  Gain insights and detailed data on the computing industry, including salary, job outlook, 'day in the life' videos, education, and more with our new Career Insights.
-  Redesigned job search page allows you to view jobs with improved search filtering such as salary, location radius searching and more without ever having to leave the search results.
-  Receive the latest jobs delivered straight to your inbox with **new exclusive Job Flash™ emails**.
-  Get a free resume review from an expert writer listing your strengths, weaknesses, and suggestions to give you the best chance of landing an interview.
-  Receive an alert every time a job becomes available that matches your personal profile, skills, interests, and preferred location(s).

Visit <https://jobs.acm.org/>

CAREERS

Centre College

Tenure Track Assistant Professor of Computer Science

Centre College invites applications for a tenure track position in Computer Science beginning August 2022. Candidates must demonstrate a strong commitment to teaching in a student-centered environment. Special consideration will be given to candidates committed to using/developing inclusive pedagogies, mentoring student research, and supporting students beyond the classroom. The successful candidate must have a Ph.D. or equivalent terminal degree by the start of their appointment. This is an expansion position in a growing program that offers majors and minors in both Computer Science and Data Science. Candidates with all specializations are encouraged to apply. More information about the Computer Science program can be found here: <https://www.centre.edu/majors-minors/computer-science/>.

To apply, please go to <http://apply.interfolio.com/91004>. A completed application will include: 1) letter of application, 2) curriculum vitae, 3) a statement of teaching philosophy, experience, and effectiveness, 4) a diversity statement explaining how the candidate would contribute to and/or address the issues of diversity and inclusion at Centre, 5) transcripts, and 6) three confidential letters of reference from individuals able to share insight into the candidate's professional expertise and future potential. Review of applications will begin October 10, 2021, and continue until the position is filled.

Established in 1819, Centre College has long been ranked among the top fifty National Liberal Arts Colleges by U.S. News and World Report and is proud to host one of the nation's premiere study abroad programs. The Centre Commitment guarantees that all students can study abroad, have an internship or research opportunity, and graduate in four years. With 1,400 students and an exceptional faculty of teacher-scholars, classes are small and academic standards are high, and Centre graduates enjoy extraordinary success in top graduate and professional schools, prestigious fellowships, and rewarding careers.

Centre College is committed to an environment that welcomes and supports diversity. As noted in the Statement of Community, Centre strives to create an environment where differences are celebrated, governance is shared, ideas are freely and respectfully exchanged, and all members of the community benefit from the richness of diverse backgrounds and experiences. Therefore, the Computer Science program strongly encourages applications from candidates who further diversify our faculty, who celebrate the rich diversity of our student body, and who utilize inclusive and engaging pedagogical practices. A number of resources support faculty success, including a robust Center for Teaching and Learning, peer mentoring, membership in the National

Center for Faculty Development and Diversity, and extended funding for professional development.

Centre is located in Danville, Kentucky, a city of 18,000 recognized for its high quality of life, historic downtown, friendly people, beautiful landscapes, and easy access to Lexington, Louisville, and Cincinnati. For more information about Centre College, visit our website at www.centre.edu.

Stanford University

Faculty Positions in Operations, Information and Technology

The Operations, Information and Technology (OIT) area at the Graduate School of Business, Stanford University, is seeking qualified applicants for full-time, tenure-track positions, starting September 1, 2022. All ranks and relevant disciplines will be considered. Applicants are considered in all areas of Operations, Information and Technology (OIT), including the management of service and manufacturing systems, supply and transportation networks, information systems/technology, energy systems, and other systems wherein people interact with technology, markets, and the environment. Applicants are expected to have rigorous training in management science, operations research, engineering, computer science, economics, and/or statistical modeling methodologies. Candidates with strong empirical training in economics, behavioral science or computer science are encouraged to apply. The appointed will be expected to do innovative research in the OIT

field, to participate in the school's PhD program, and to teach both required and elective courses in the MBA program. Junior applicants should have or expect to complete a PhD by September 1, 2022.

Applicants should submit their applications electronically by visiting the web site <http://www.gsb.stanford.edu/recruiting> and uploading their curriculum vitae, research papers and publications, and teaching evaluations, if applicable, on that site. Applications will be accepted until November 30, 2021. **For an application to be considered complete, the applicant must submit a CV and job market paper and arrange for three letters of recommendation to be submitted before the application deadline of November 30, 2021.**

The Stanford Graduate School of Business will not conduct interviews at the INFORMS meeting in Anaheim, but some OIT faculty members will attend. Hence candidates who will be presenting at INFORMS are encouraged to submit their CV, a research abstract, and any supporting information before October 11, 2021.

Any questions regarding the application process should be sent by email to Faculty_Recruiter@gsb.stanford.edu.

Stanford is an equal employment opportunity and affirmative action employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, protected veteran status, or any other characteristic protected by law. Stanford welcomes applications from all who would bring additional dimensions to the University's research, teaching and clinical missions.



ADVERTISING IN CAREER OPPORTUNITIES

How to Submit a Classified Line Ad: Send an e-mail to acmm mediasales@acm.org. Please include text, and indicate the issue/or issues where the ad will appear, and a contact name and number.

Estimates: An insertion order will then be e-mailed back to you. The ad will be typeset according to CACM guidelines. NO PROOFS can be sent. Classified line ads are NOT commissionable.

Deadlines: 20th of the month/2 months prior to issue date. For latest deadline info, please contact:

acmm mediasales@acm.org

Career Opportunities Online: Classified and recruitment display ads receive a free duplicate listing on our website at:

<http://jobs.acm.org>

Ads are listed for a period of 30 days.

For More Information Contact:

**ACM Media Sales
at 212-626-0686 or
acmm mediasales@acm.org**



Association for
Computing Machinery

ACM Transactions on Computing for Healthcare (HEALTH)

*A multi-disciplinary journal for
high-quality original work on how
computing is improving healthcare*

ACM Transactions on Computing for Healthcare (HEALTH) is the premier journal for the publication of high-quality original research papers, survey papers, and challenge papers that have scientific and technological results pertaining to how computing is improving healthcare.



For further information and to submit
your manuscript, visit health.acm.org

[CONTINUED FROM P. 96] defined by A and the arc from Pup to Pdown and subtract from that the area of the triangle defined by A, Pup, and Pdown. Then we do the same thing starting from B.

Suppose randomower is attached at 550 meters along the rope. Let point P be the point exactly between A and B. Consider the right triangle defined by post A, point P, and point Pup. The distance from P to Pup is $\sqrt{((550)^2 - (500)^2)}$ or $50 * \sqrt{21}$ on each side of the line segment, which is approximately 229.

The area of the triangle P, Pup, A is 57,282. This must be doubled because we must add in the triangle Pdown, Pup, A. The doubled area is 114,564.

The arc defined by A, Pup, and Pdown has a radius of 550 and an interior angle of 0.86 radians. So the area of the pie described by this portion of the circle is $(550)^2 * 0.86/2 = 130,140$ square meters. The difference of the pie area (130,140) and the triangle area (114,564) is 15,576 square meters. Double this because of the similar arc anchored at post B. **End of solution.**

Given this geometrical intuition, you are now prepared for Upstarts.

Upstart 1: If there are k attachments available, where should you put them to mow the greatest amount of area?

Upstart 2: How many attachments are needed to draw a segment of width w for some w less than 70 and where should those attachments be placed?

Upstart 3: Suppose A and B play a game in which each wants to use attachments that cut as much of the lawn as possible. They take turns as follows: A makes the first attachment. Then B makes two. Then A makes two. This goes on until each makes the same number of moves (A determines when to make last move by making a single attachment). Each player gets credit for every part of the lawn that is first mowed by randomower due to an attachment by that player. Please find a winning strategy for some player.

Dennis Shasha (dennisshasha@yahoo.com) is a professor of computer science in the computer science department of the Courant Institute at New York University, New York, NY, USA, as well as the chronicler of his good friend the omniheurist Dr. Ecco.

All are invited to submit their solutions to upstartpuzzles@cacm.acm.org; solutions to upstarts and discussion will be posted at <http://cs.nyu.edu/cs/faculty/shasha/papers/cacmpuzzles.html>

Copyright held by author.



Dennis Shasha

DOI:10.1145/3477385

Upstart Puzzles

Randomower

End-of-the-rope machinations.

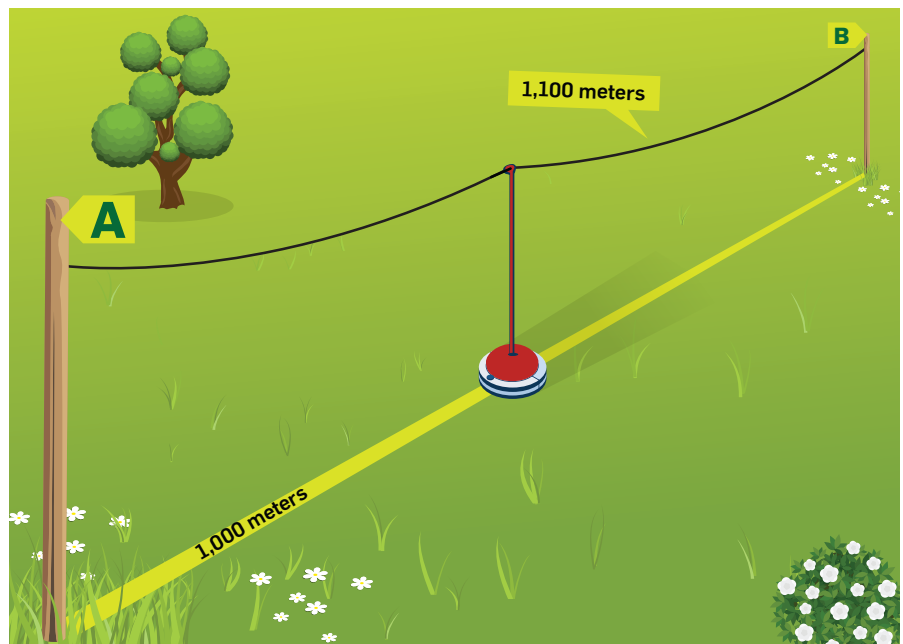
THERE IS ONE perfectly flexible but unbreakable rope attached to two posts on a large but unkempt lawn. The posts are 1,000 meters apart and the rope is 1,100 meters long. So there is some slack in the rope.

A randomower is a driverless lawnmower that moves randomly and is able to change direction by an arbitrary angle at any time. Fortunately, it can be clipped to the rope so it will not wander off too far. You are trying to use this random lawnmower to cut at least part of the grass between the posts. Let the rope's zero point be at post A and its end point (1,100 meters away along the rope) at post B.

Suppose you wanted to cut the grass to clear a path, that is, a line segment at least the width of the lawnmower itself, between the two posts. It is OK if more grass is cut to the side of the path, but you want to ensure you have a straight continuous path. The first challenge is to do this with the minimum number of attachments.

Warm-up: What is the smallest number of attachments necessary to mow a straight line segment (and, optionally, other grass) from post to post?

Solution to Warm-Up: Attach the randomower to the 100-meter mark along the rope starting say from post A. Because the rope is 100 meters longer than the distance between the posts, the distance along the rope between the attachment point of the randomower and post B is 1,000 meters. Therefore, the randomower can reach post A up to a point 100 meters from post A in the direction of B. Next, attach the randomower at the 200-meter mark along the rope from post A. Keep



What is the maximum area this randomly moving lawnmower (randomower) can mow, while staying attached at a single point on the rope?

going for 300, 400, ... 1,000. So, all together we will need 10 attachments.

Now, let's reset the lawn to its unkempt state and ask a few other questions. If we were to attach the randomower to a single point along the rope, where should we attach it to cut the greatest area of grass? It might seem that near an end might be best, though symmetry suggests the middle.

Warm-up 2: Given some attachment point $p = 400$ meters from post A, what would be the maximum width of the mowed area?

Solution to Warm-up 2: If $p = 400$, we have a triangle of length $p = 400$, $1,100 - p$, and $1,000$. This gives us an interior angle of 0.31756 rad (thank you, Wolfram Alpha). So the width will be $700 \cdot$

$\sin 0.31756 = 218$. That is just one side, so we must double this to get the total width. Notice this is a lot more than 99.8 meters, which is what we would get by attaching to $p = 100$.

So, now we can ask where would be the best place to attach randomower to achieve the largest area.

Challenge: What would be the area of the mowed grass if you attached the randomower to the 550-meter point along the rope?

Solution: Denote by Pup and Pdown the most extreme points on either side of the line segment between A and B when attaching randomower to the 550-meter point along the rope. To compute the area, we must consider the pie-shaped area [CONTINUED ON P. 95]



ACM BOOKS

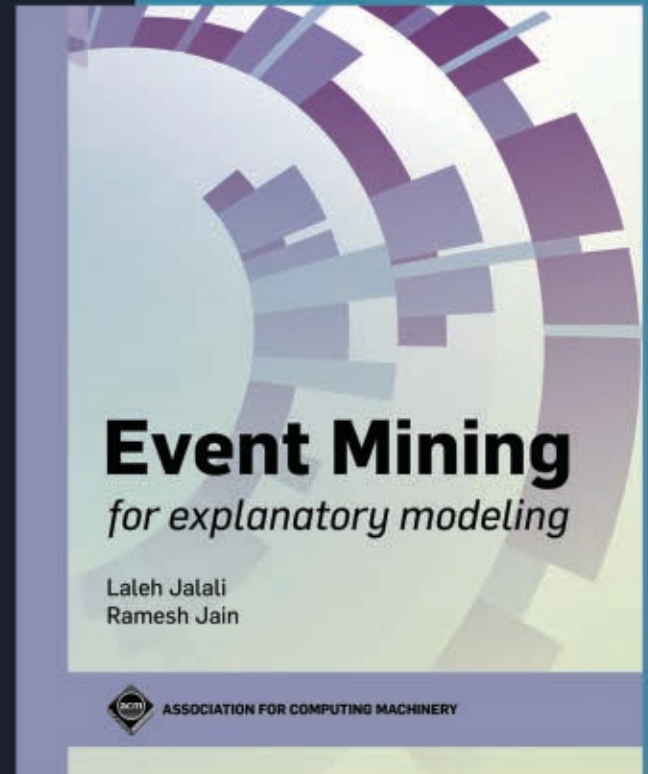
Collection II

This book introduces the concept of Event Mining for building explanatory models from analyses of correlated data. Such a model may be used as the basis for predictions and corrective actions. The idea is to create, via an iterative process, a model that explains causal relationships in the form of structural and temporal patterns in the data. The first phase is the data-driven process of hypothesis formation, requiring the analysis of large amounts of data to find strong candidate hypotheses. The second phase is hypothesis testing, wherein a domain expert's knowledge and judgment is used to test and modify the candidate hypotheses.

The book is intended as a primer on Event Mining for data-enthusiasts and information professionals interested in employing these event-based data analysis techniques in diverse applications. The reader is introduced to frameworks for temporal knowledge representation and reasoning, as well as temporal data mining and pattern discovery. Also discussed are the design principles of event mining systems. The approach is reinforced by the presentation of an event mining system called EventMiner, a computational framework for building explanatory models. The book contains case studies of using EventMiner in asthma risk management and an architecture for the objective self. The text can be used by researchers interested in harnessing the value of heterogeneous big data for designing explanatory event-based models in diverse application areas such as healthcare, biological data analytics, predictive maintenance of systems, computer networks, and business intelligence.

<http://books.acm.org>

<http://store.morganclaypool.com/acm>



Event Mining *for explanatory modeling*

Laleh Jalali
Ramesh Jain

ISBN: 978-1-4503-8483-4

DOI: 10.1145/3462257



MATLAB SPEAKS MACHINE LEARNING

With MATLAB® you can use clustering, regression, classification, and deep learning to build predictive models and put them into production.

mathworks.com/machinelearning

©2021 The MathWorks, Inc.