Check for updates

BLOG@CACM

The *Communications* website, https://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.

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

DOI:10.1145/3639711 https://cacm.acm.org/blogs/blog-cacm

Pondering the Ugly Underbelly, and Whether Images Are Real

Robin K. Hill teaches how proofs lead to the truth, while Carlos Baquero searches for truth in imagery.

Robin K. Hill Cook-Levin: The Ugly Underbelly is Good for Us

https://bit.ly/3RFrcUC
December 10, 2023

I tell my computer science students that we go through proofs not to show *what* is true, but *why* it is true. During an interesting consultation with a graduate student who wanted to know whether I proved the Cook-Levin Theorem when I taught that result in the Foundations of Computing class, I said I had, indeed, and every teacher should. As is the human predilection, I became ever more fully convinced I was right as I expounded on it.

The textbook in use is Michael Sipser 3rd Edition.[1] In Section 7.4, we see Cook-Levin as Theorem 7.37: SAT is NP-Complete. The easy part is showing that SAT is in NP, but the gloriously laborious step, the nitty-gritty, is showing every language A in NP is polynomially time reducible to SAT. This involves pages of close reasoning about the Turing Machine (TM) N that decides A.

What have we got to work with—a language A in NP, and therefore, a nondeterministic TM N that decides A in $O(n^k)$ time for some k and input w of length n, and therefore, some accepting computation history of the TM for each word w in A. What are we after?—a propositional formula ϕ that has a satisfying assignment of truth values if and only if N accepts w.

We create a tableau. We hypothesize an arbitrary word $w \in A$, and plant

What are we after? A propositional formula ϕ that has a satisfying assignment of truth values if and only if N accepts w.

an accepting branch of its computation history (configurations) in rows of that tableau. We subscript each cell of the table, and form variables from the subscripts and the symbols (the components of N) that will participate in the formula ϕ, which consists of four conjuncts, each derived with persnickety care from the TM on w. The first establishes the symbol contents of the tape cells, given by disjunct clauses of enumerations of the possibilities, and their settings (true or false) by a conjunct of legality clauses. The other main conjuncts comprise one establishing that the first configuration contains the start state, one doing similar for an accepting state, and one verifying the legality of each move according to the TM's transitions. For that last, we define a window, a frame of six before-and-after tape cells, constructed in excruciating detail, that can be compared to the possible legal transitions of N. And so on ... These four conjuncts come together as the ϕ. Important and subtle point for students: The fact that ϕ has a satisfying assignment is

known by the construction of the first conjunct, although the satisfying assignment itself cannot be constructed since we do not know N or w.

How painful is this? ... is a pointless question. We have arrived at the SAT formula ϕ. Now we calculate the complexity measure of each part, each shown to be polynomial, with their sum therefore also polynomial. Therefore, A ∈ NP. Important and subtle point for students: We have shown this for *every* language A in NP, by ignoring any particulars of A and ignoring any particulars of the string w, that is, without loss of generality.

Why should we go to all this trouble, and drag our students through it? Because it shows, indispensably, profoundly, and distinctively, the nature of computability. Here is a TM, which consists of sets (finite!), supplying givens (definitions!) that provide the (exact!) tools necessary to reach the conclusion. That's all there is. Proofs of other theorems can unveil the same picture, cultivate the same appreciation of the TM concept, but this proof of Cook-Levin seems particularly rich in its layers of exposition, drawing newcomers in so deeply that they inhabit the result. (And it could be even more complicated: Notwithstanding the excruciating detail, Sipser forgoes a precise definition of a legal window.) At all points along the way, we observe where we came from, where we're going, the path we're following, and the vehicle we're taking.

There is no essence of a TM; there is only the TM itself, in its five-part or seven-part glory (a 7-tuple in Sipser), and in the protocol for its execution, which applies to all TMs in the class induced by the definition. There is no abstraction or generalization. We humans extrapolate and profit from hand-waving, from simplification and analogy, the type of explanation that I have given verbally. But there is no substitute for the degree of detail required to turn a TM into an SAT formula.

There is no more: There is no result beyond that which can be obtained from using the symbols, states, and transitions on some input string. A Turing Machine proffers no interpretation. There is no less: There is no shortcut, no gloss. Every bit of the components matters—each state, each symbol from the alphabets, and each transition. A Turing Machine proffers no summary.

Reference

1. Sipser, M. *Introduction to the Theory of Computation.* Cengage Learning. 2012.

Carlos Baquero
All Photos Are Fake Until Proven Real

https://bit.ly/3vnlXRZ

November 1, 2023

When we look at a painting, we rarely expect it to depict reality with fidelity; we look for some meaning, aesthetic value, and possibly a message or a lasting feeling. With photography, we still expect them to depict a snapshot of reality taken at a given space and time. This is changing, and fast (https://politi.co/3RD1wb2).

Conflicts are great motivators for the manipulation of information. Genuine photos have been labelled as fakes, after false positives in artificial intelligence (AI) image-detection tools. Fake photos, with clear AI telltale signs (for example, hands with seven fingers) have been disseminated as genuine. Fake photography is now easily accessible and of increasing quality, making it easy to generate images like the ones below, done in a couple of minutes on a smartphone with DALL.E 3 (an AI image generator from prompts, by OpenAI). Soon these images, here a fake tornado event and Eve tempted by the forbidden fruit, will become indistinguishable from real photography, and maybe help fuel misinformation.

We are quite able to distinguish fiction from reality when we go and watch a movie or when playing with a VR headset. But the mix of real and fake images in news and advertising will erode our trust in digital information. Recently, photographer Boris Eldagsen won, then declined a Sony World Photography Award with an AI image. The line between *Promptography* and traditional *Photography* is becoming increasingly blurred.

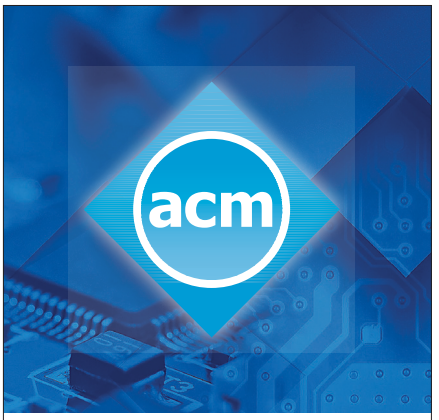
Conflicts are great motivators for the manipulation of information.

Top, a tornado in a desert. Bottom, a person holding an apple.

What Can Be Done?

The first step is to rewire our perception and start assuming by default that photographic content (video as well) is not necessarily grounded in reality. Let us draw a parallel with official documents. Anyone can produce a PDF document that looks like a Ph.D. certificate, but that document has no value unless signed by a reputable institution that can issue doctorate degrees. The value of the content depends not only on the content, but on who signs for the content. Likewise, the same scientific paper in a pre-print archive and in a good journal with quality peer review can have the same content, but a very different value. Since photography used

IMAGES BY CARLOS BAQUERO (USING DALL.E 3)



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
acmmediasales@acm.org

In the absence of watermarking, it is tempting to resort to tools that try to separate AI content from genuine content.

to depict reality, or at least a close approximation of reality, we have been neglecting the certification component.

Watermarking and Detection

In a statement from July 2023, the Biden Administration reported on voluntary commitments from major AI companies on various aspects of responsible AI. These include "developing robust technical mechanisms to ensure that users know when content is AI-generated, such as a watermarking system." Watermarking technology can annotate images without apparent degradation of quality. Google's SynthID tool uses neural networks to add and check watermarks, and announces that they can even be detected after intermediate screenshots, image rotations, and resizing.

Watermarking, once adopted by the major providers of AI images, will help pinpoint images generated at the request of average users. However, sophisticated users can run their own personal AI imaging platforms and avoid watermarking. Also, the source of disinformation campaigns often involves adversary state actors that can easily run their own high-grade platforms.

In the absence of watermarks, it is tempting to resort to tools that try to separate AI content from genuine content. However, these tools are subject to misclassification both by false positives and false negatives. Lock manufacturers and lockpickers have been in an arms race for centuries; we can expect a similar arms race between AI generation and detection algorithms.

Content Provenance and Authenticity

An alternative to the quest to identify artificial content is to provide tools for certifying genuine images and videos. The idea is to have dedicated software and hardware in the camera that cryptographically signs the image content and context information. Further signatures can be added as the image is processed and finally published to the public. Each step adds another link to the chain of provenance. The Coalition for Content Provenance and Authenticity (C2PA, https://c2pa.org/) provides tools and technical standards for certifying the source and history of media content. This includes an icon on the image that, when inspected, provides provenance metadata to the final consumer.

Long before the advent of AI imaging, disinformation tactics frequently involved misattributing authentic images from one event to a different time and place. Fortunately, provenance can counteract such traditional manipulations. Users already have the ability to verify digital certificates, ensuring secure website interactions and validating signatures in PDF documents. While adopting new practices can be challenging, as this technology becomes more prevalent, users will possess a potent tool to verify content authenticity as they consume it.

As always, the most essential tool is not technological but the readers' exercise of their own critical thinking. Before accepting recent events, it is crucial to corroborate them with multiple sources. As Carl Sagan aptly said, "Extraordinary claims require extraordinary evidence."

Acknowledgments

I would like to thank António Coelho for comments on improving this text.

Robin K. Hill is a lecturer in the Department of Computer Science and an affiliate of both the Department of Philosophy and Religious Studies and the Wyoming Institute for Humanities Research at the University of Wyoming, Laramy, WY, USA. **Carlos Baquero** is a professor in the Department of Informatics Engineering within the Faculty of Engineering at Portugal's Porto University and also is affiliated with INESC TEC. His research is focused on distributed systems and algorithms.

© 2024 Copyright held by the owner/author(s).