AI and Digital Forensics

Dr Simson Garfinkel, Chief Scientist
BasisTech
simsong@basistech.com

NIST Long-Term Vision and Strategic Priorities for Forensic Science in the United States: Roundtable Discussion with Thought Leaders
Sept. 6-7, 2023

BasisTech

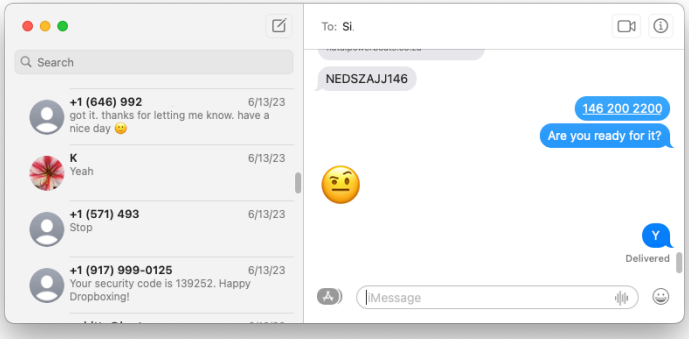
AI and Digital Forensics: Two Possible Meanings

AI for Digital Forensics

Helping forensic practitioners

Interpreting evidence

Performing investigations



Digital Forensics for AI Systems

Explaining an AI decision (XAI)

Analyzing an AI system — for example, at an accident or crime scene

https://www.ntsb.gov/investigations/Pages/HWY19FH008.aspx

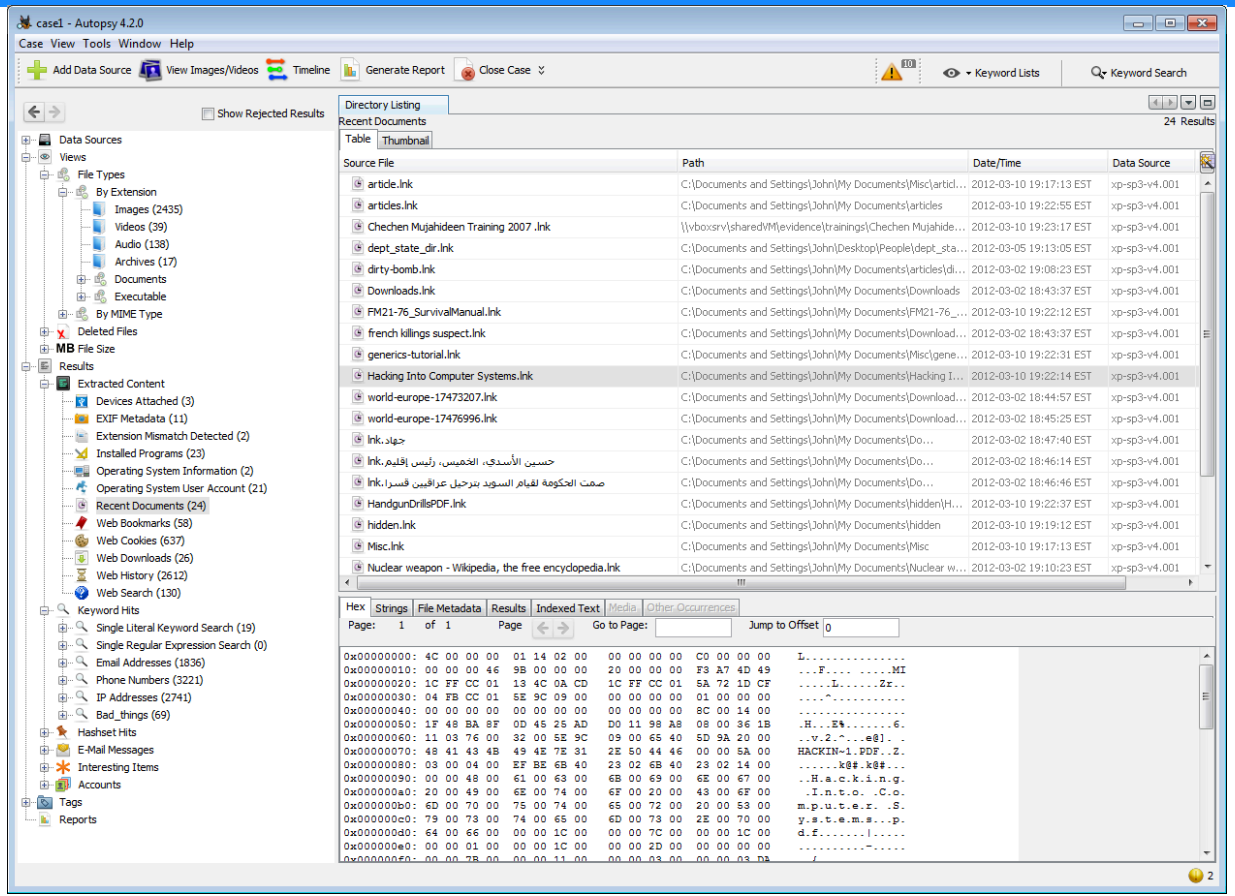
AI for Digital Forensics

AI for Digital Forensics

Let’s consider how AI might be used in:

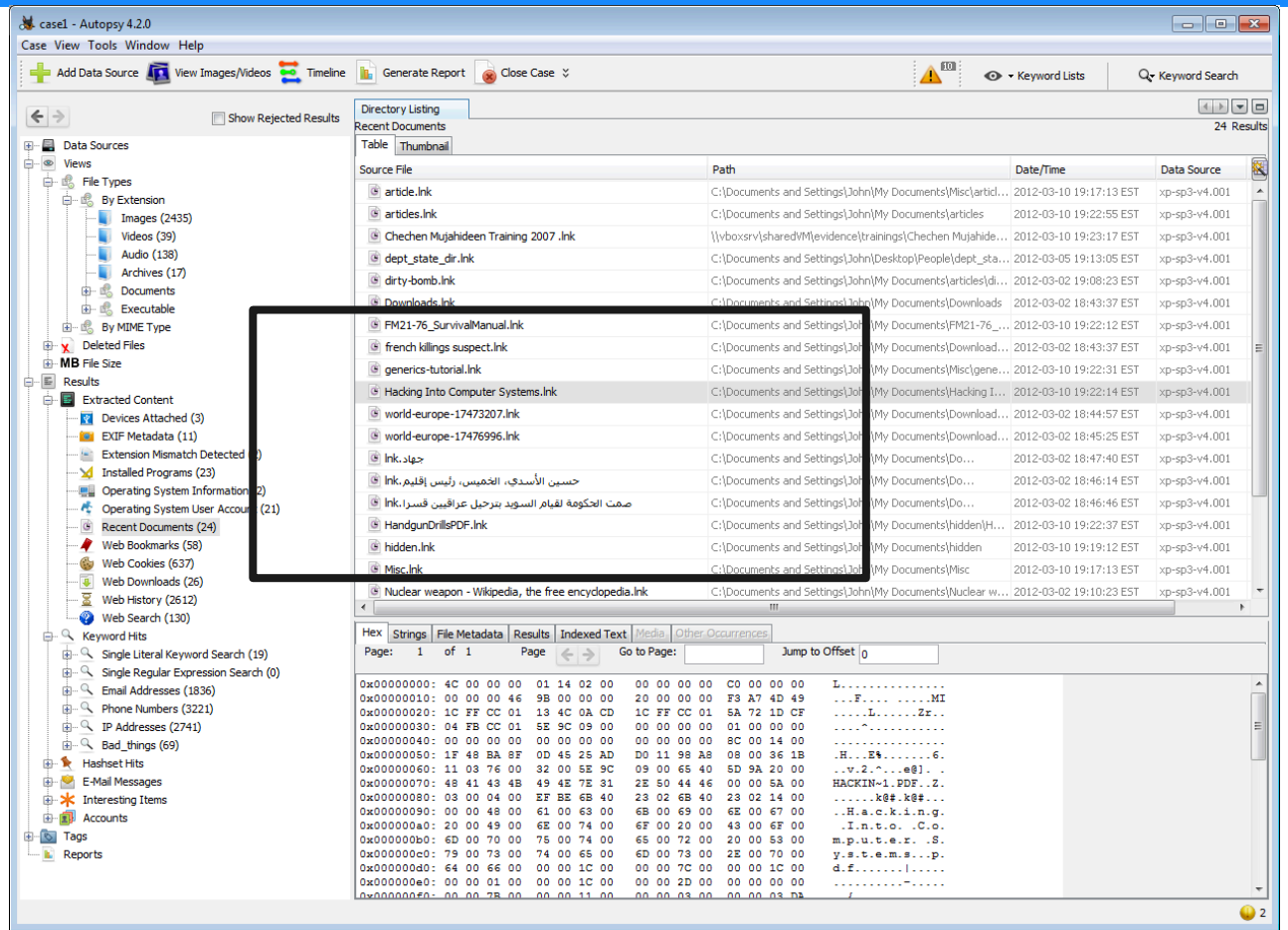
- Translation (human text)
- Translation (machine text)
- Tool recommendation
- Report writing

Translation (human text) is a traditional AI task...

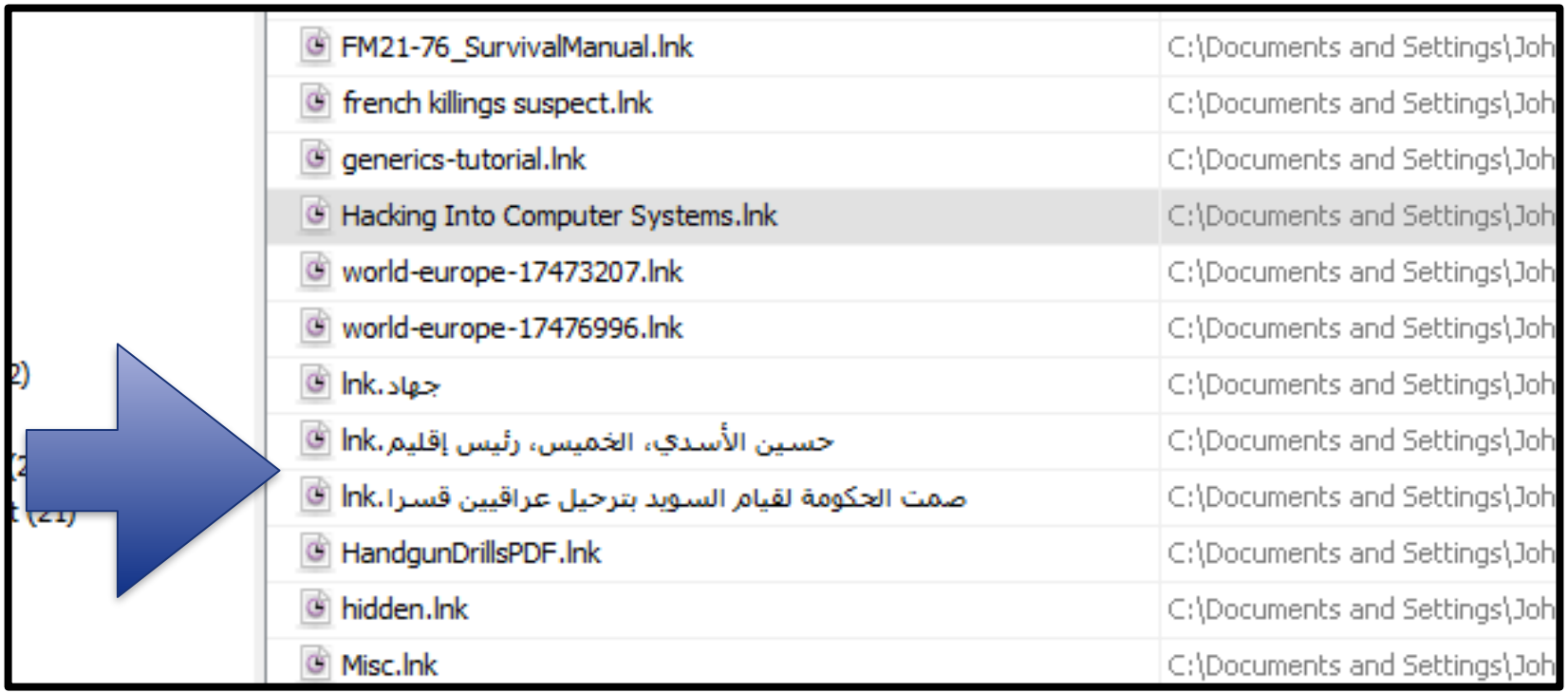


Autopsy open source digital forensics platform

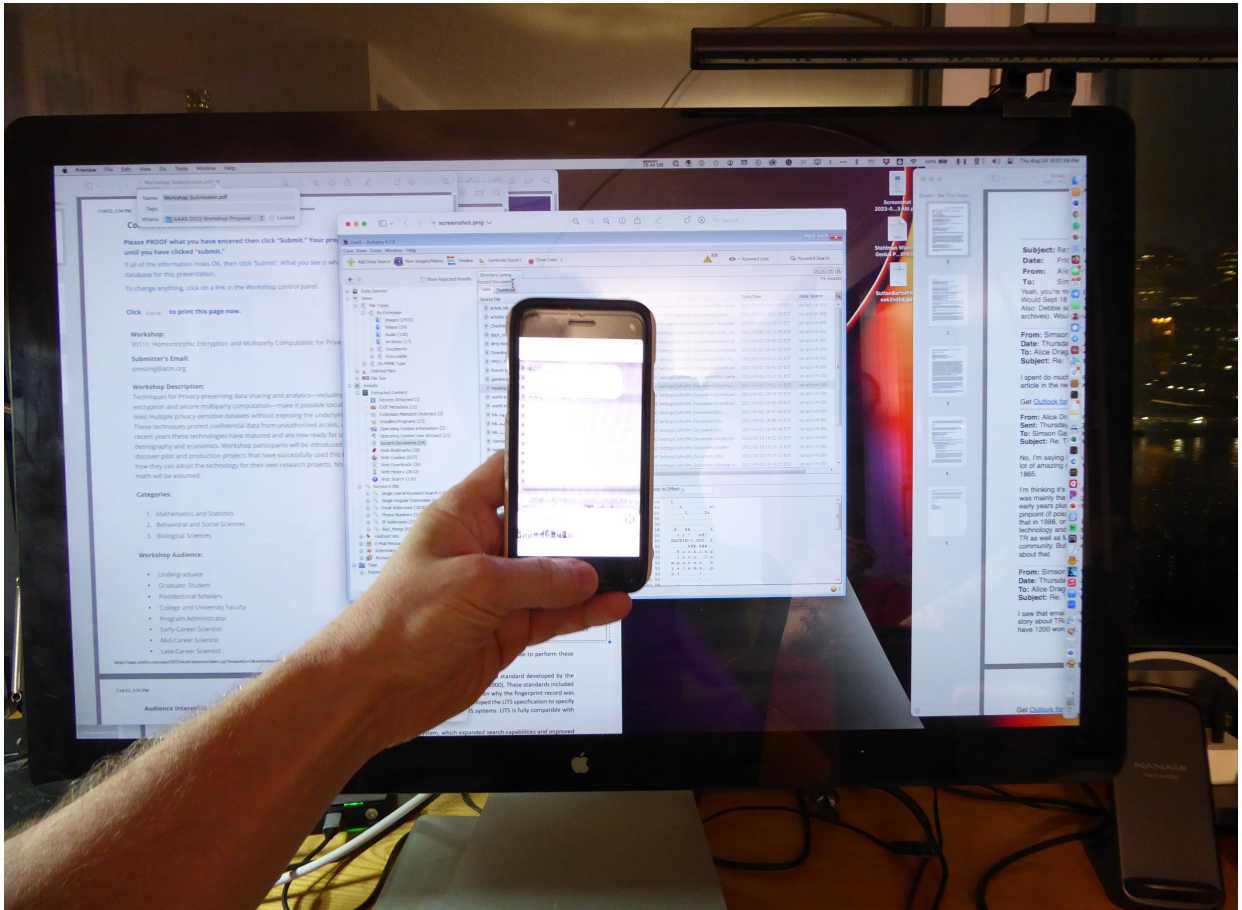
Translation (human text) is a traditional AI task...



What do these filenames mean in English?



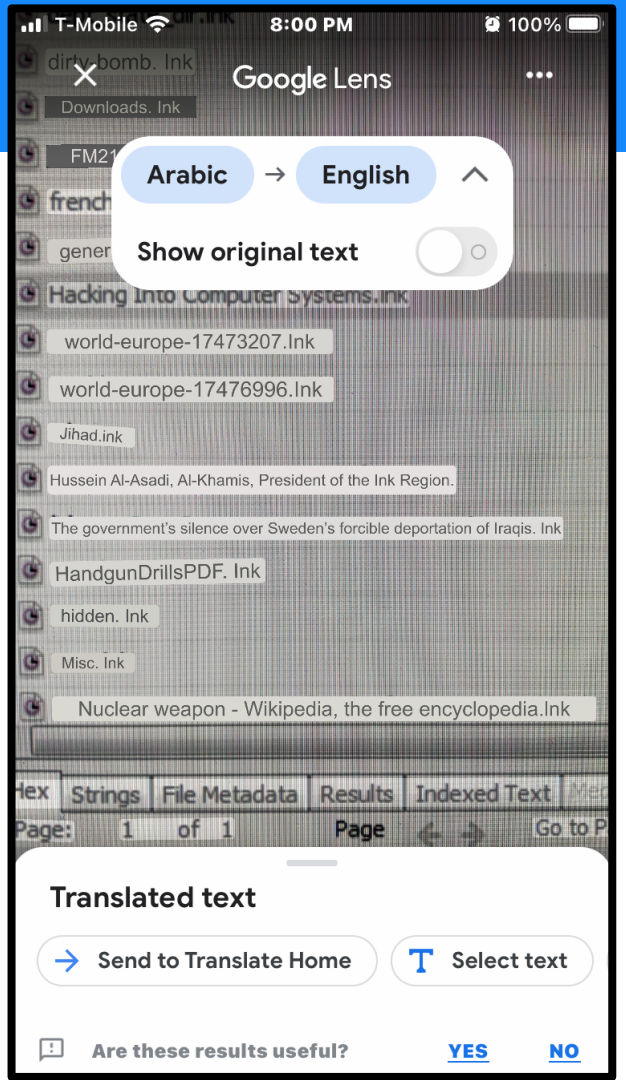
Just hold your phone up to the computer screen and use the Google Translate app...



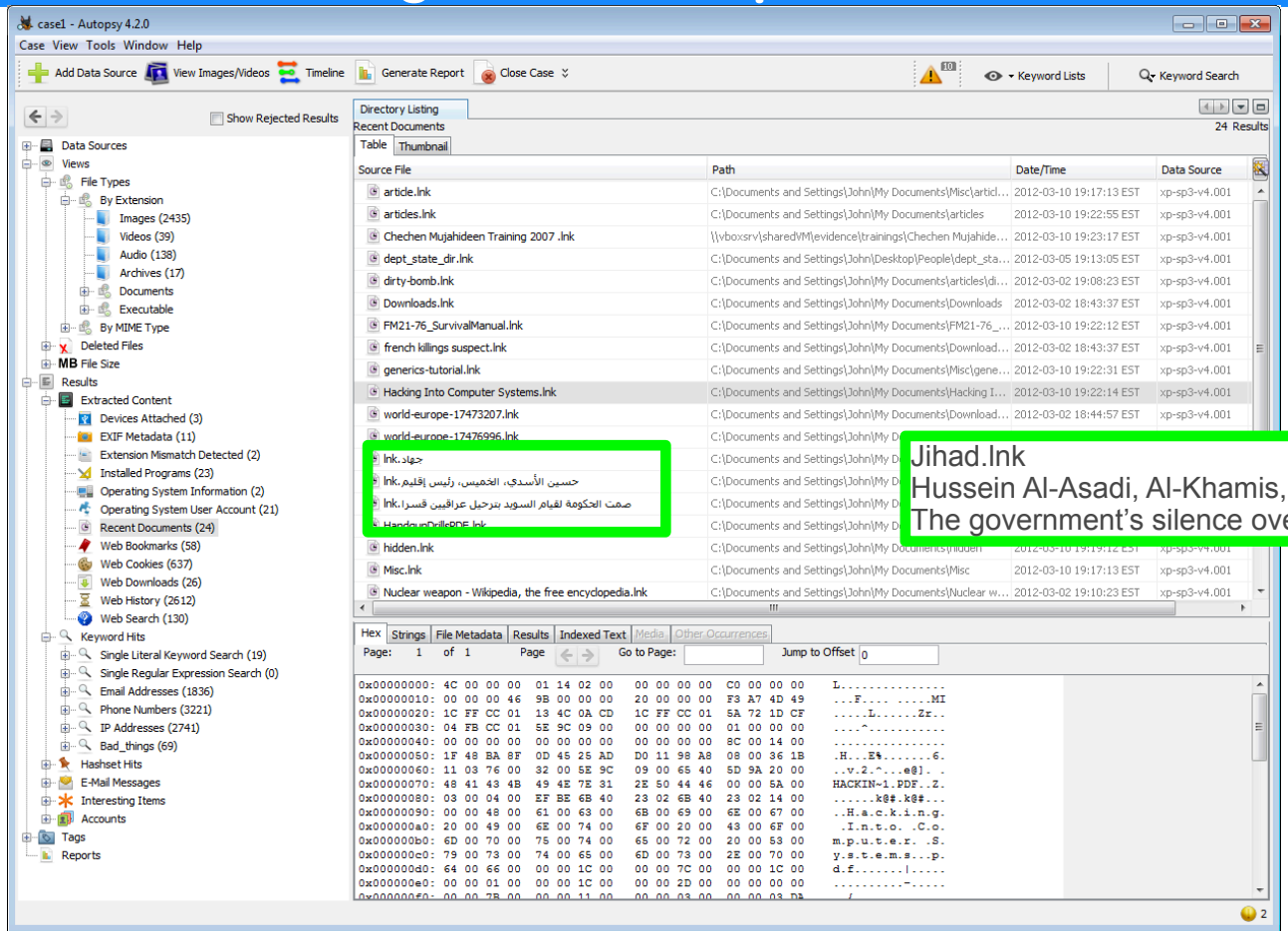
This works surprisingly well…
but it's awkward! (and hard to take into court)

Options:

- Build translation into every forensic tool
- Build translation into the OS
  - Web browsers do this already!
  - Do we need "validated translation" for forensics?



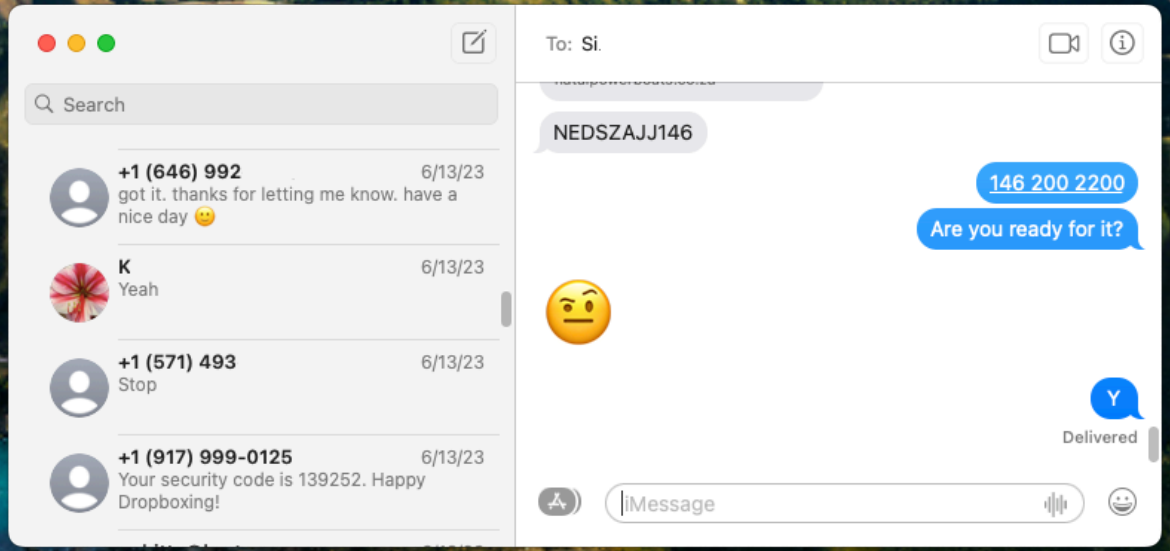
We can't update every tool — but a "forensic CoPilot" could watch the screen, look for non-English text, and provide translation in overlays…



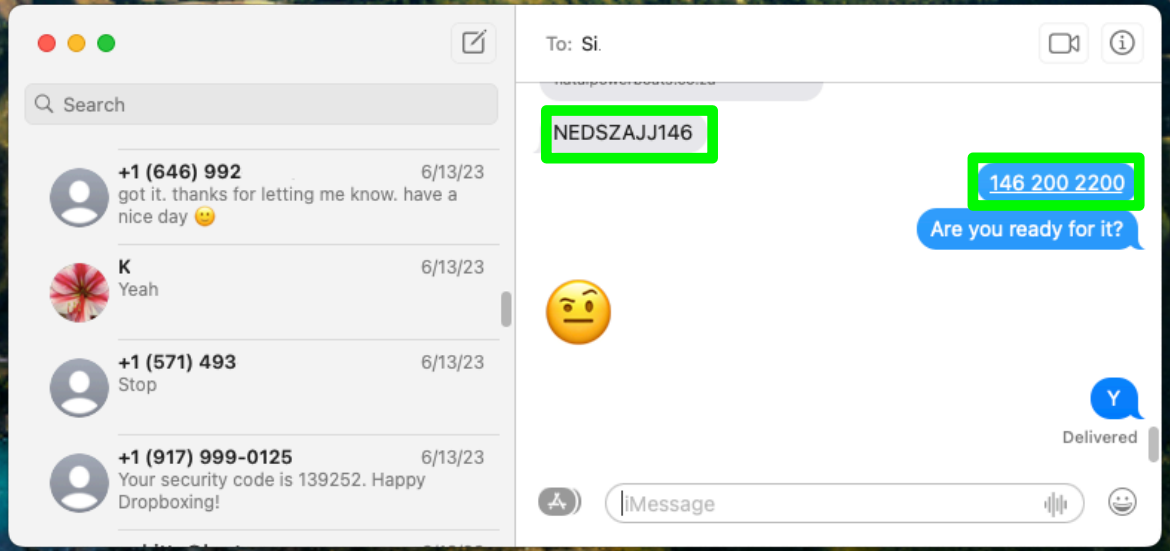
Jihad.lnk
Hussein Al-Asadi, Al-Khamis, President of the Region.lnk
The government's silence over Sweden's forcible deportation of Iraqis.lnk

The Forensic CoPilot could provide additional context. Consider a hypothetical case involving Apple Messenger...

The examiner encounters information that is unfamiliar...



AI can watch the investigator's screen and provide context.



SWIFT Code: NEDSZAJJ
NEDS - Nedbank Group LTD
ZA - South Africa
J - Johannesburg
J - (active)
146 - Branch Code
more info

146 200 2200 - Account #
Owner: Fictional Group Ltd.
more info

OSINT CoPilot

(Note: OSINT CoPilot is a hypothetical investigative tool.)

AI can recommend tools for cyber investigations.

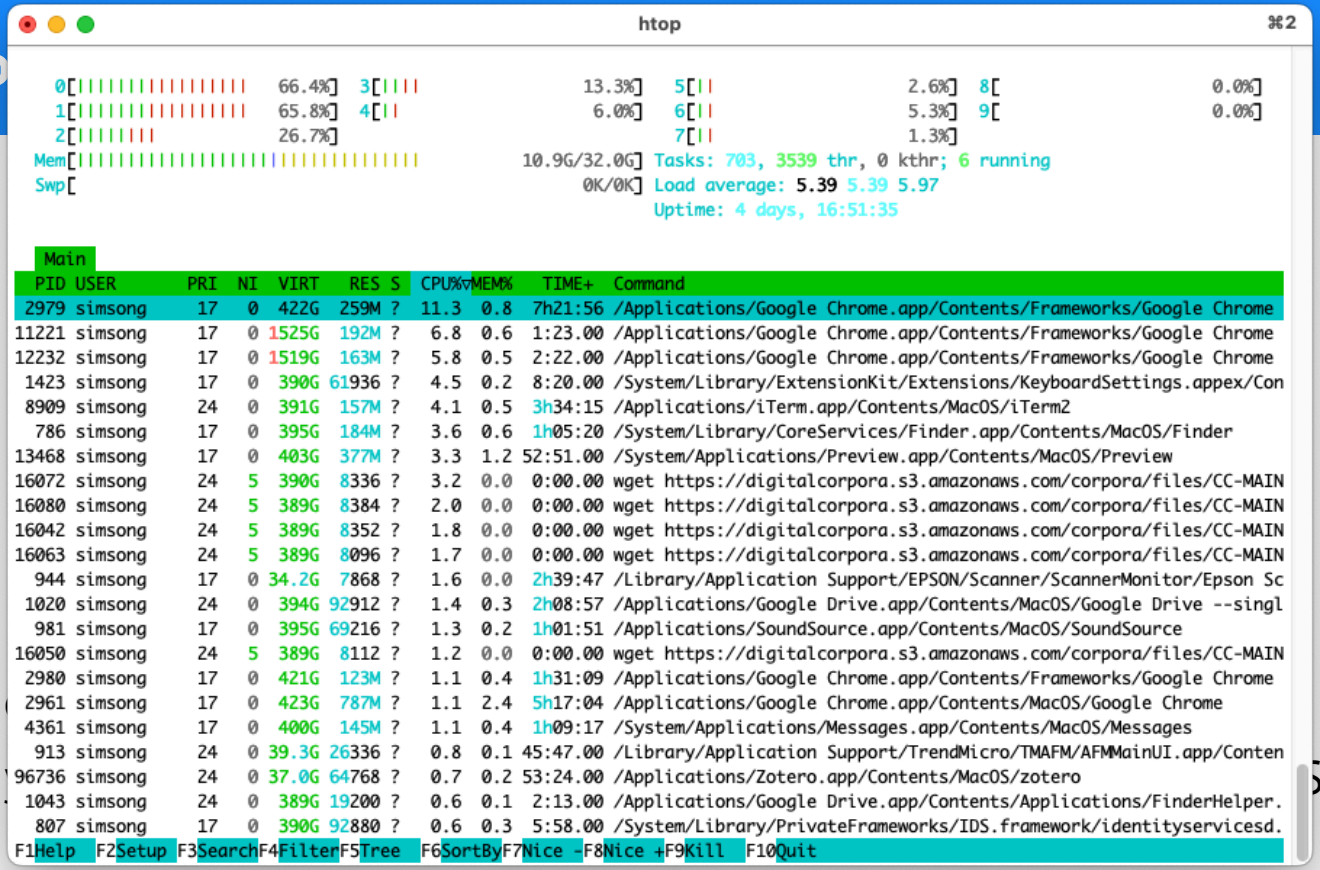
me: system sluggish.

AI: Are you on the (C) Console or accessing via (S) SSH?

me: S

AI: try: **htop**

AI can reco



me: syst

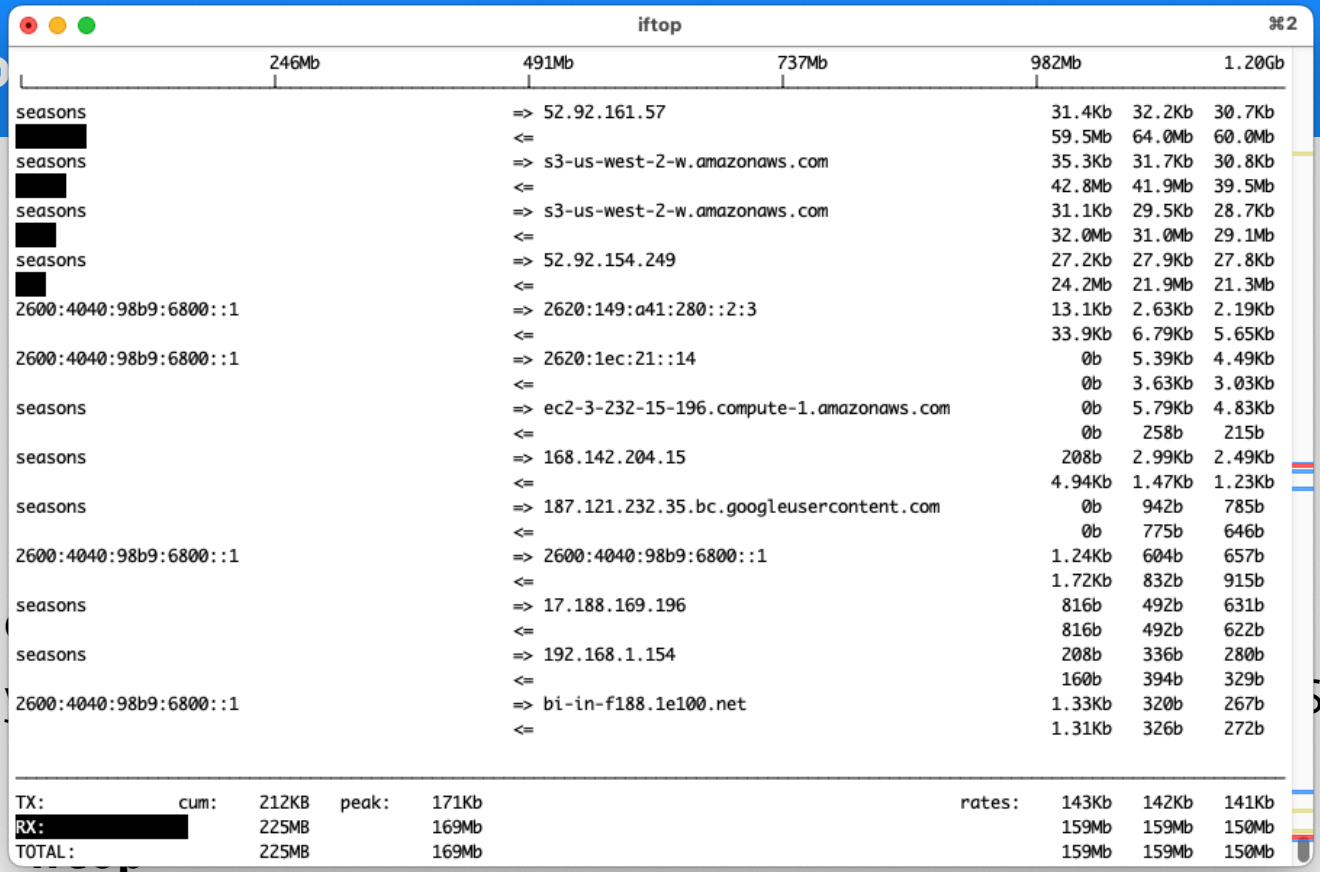
AI: Are SH?

me: S

AI: try: **htop**

AI: Sluggishness may be Chrome or wget; try: **iftop**

AI can reco



me: syst

AI: Are

me: S

AI: try:

AI: Sluggishness may be Chrome or wget; try: **iftop**

AI: … Checking IP addresses …

AI can recommend tools for cyber i

me: system sluggish.

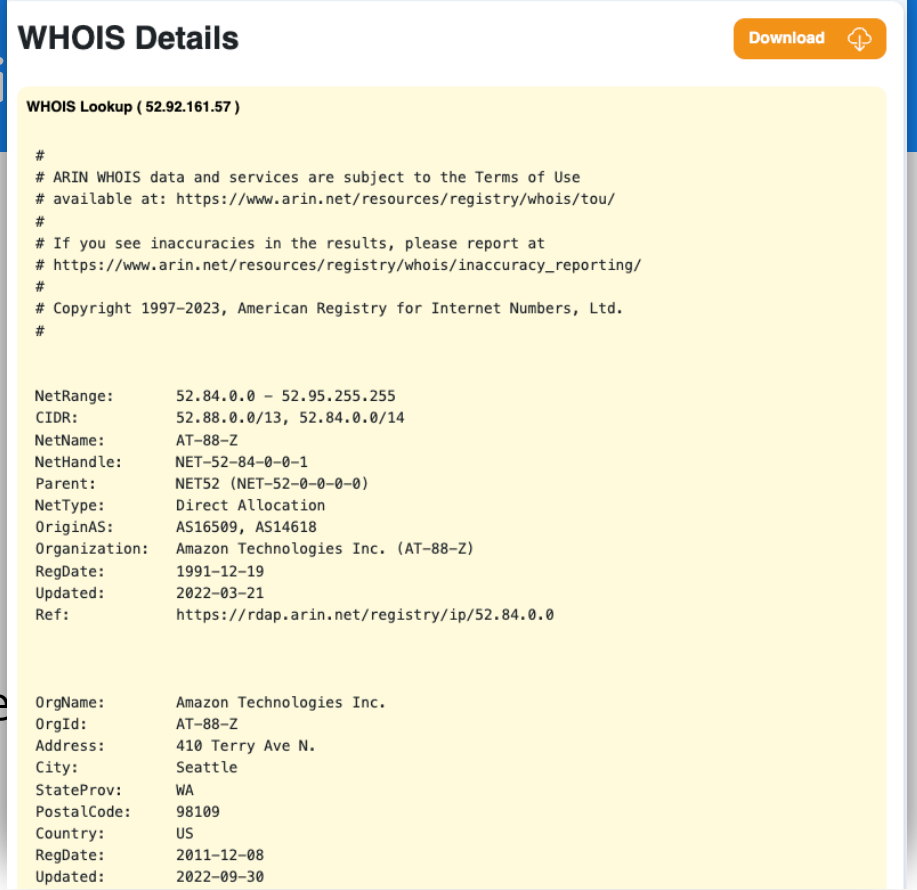
AI: Are you on the (C) Console

me: S

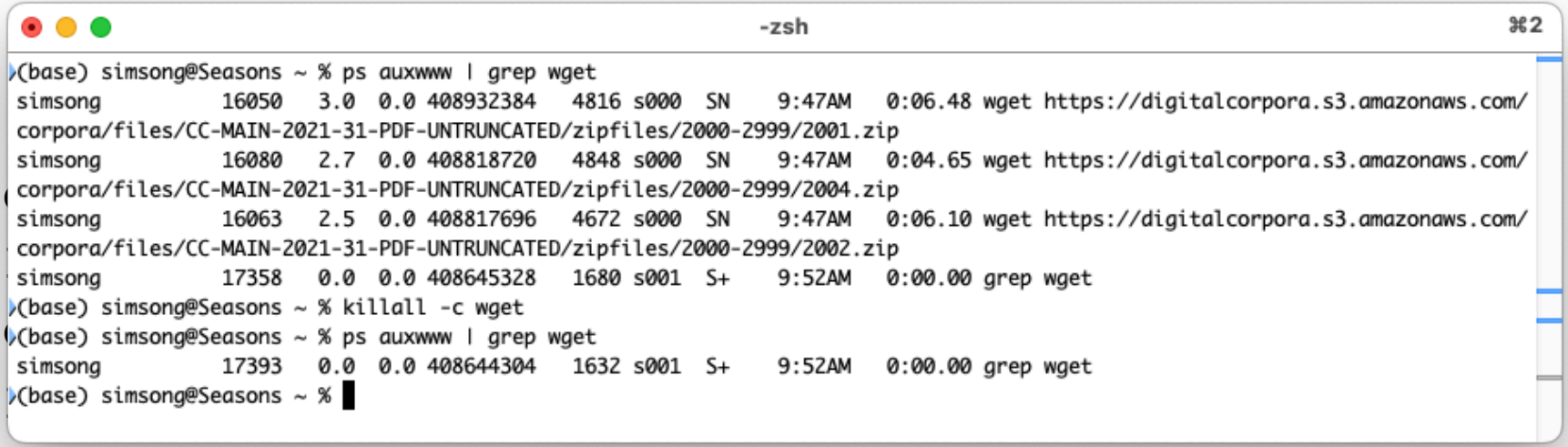
AI: try: **htop**

AI: … Checking IP addresses …

AI: Major data flows are from Amazon Web Services via wget



AI can recommend tools for cyber investigations.



AI: … Checking IP addresses …

AI: Major data flows are from Amazon Web Services via wget

AI: Recommended action: document and terminate wget processes.

AI for Digital Forensics

What I promised:

- Translation (human text) ✔
- Translation (machine text) ✔
- Tool recommendation ✔

I didn’t get to this:

- Report writing ❌
  - Didn’t get to this!

Sorry!

Please ask ChatGPT for ideas!!!

Digital Forensics for AI

Forensics will be used to bring data from AI failures into court.

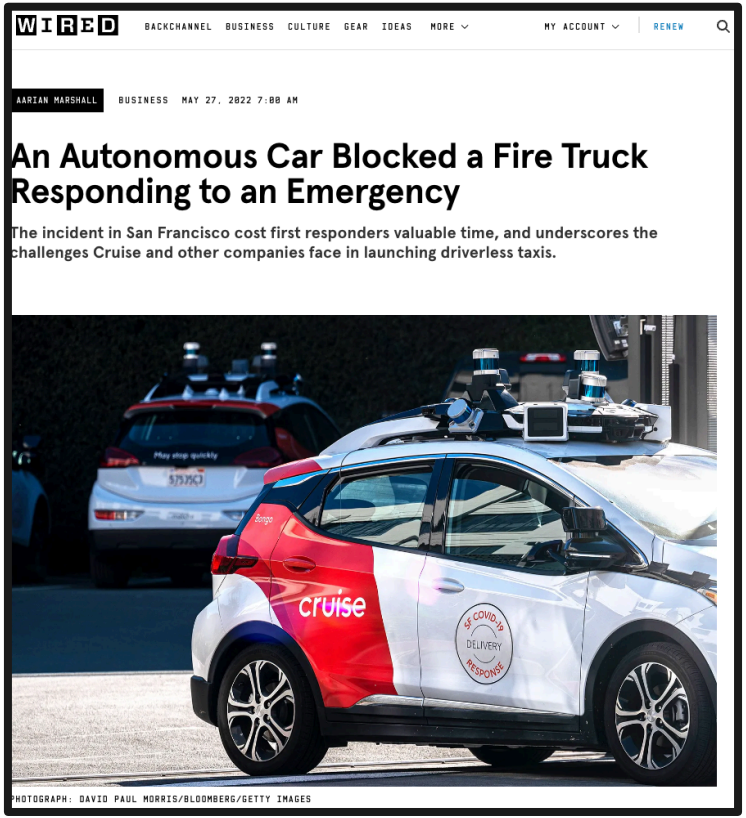
Why did the AI fail?

- Was the failure in *this particular system?*
- Was the failure in *the system's design?*

What was the source of this particular incident?

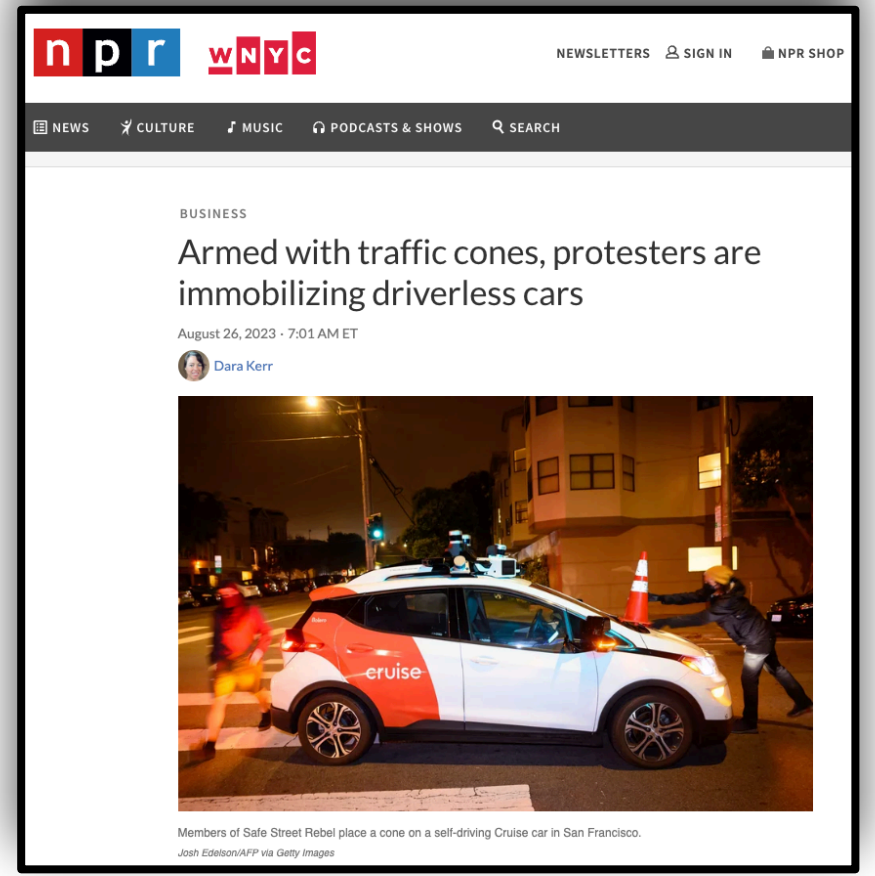
- Hardware design
- Hardware failure (sensor; computer; network)
- Software design
- Software error
- Training data coverage or bias
- A specific piece of training data
- Over-the-air update to software or model?

What kind of forensic investigation protocol would reliably answer these questions?



https://www.wired.com/story/cruise-fire-truck-block-san-francisco-autonomous-vehicles/

If attacks on driverless cars become more aggressive, forensics will be used for assessing fault



npr WNYC

NEWSLETTERS SIGN IN NPR SHOP

NEWS CULTURE MUSIC PODCASTS & SHOWS SEARCH

BUSINESS

Armed with traffic cones, protesters are immobilizing driverless cars

August 26, 2023 · 7:01 AM ET

Dara Kerr

Members of Safe Street Rebel place a cone on a self-driving Cruise car in San Francisco.

Josh Edelson/AFP via Getty Images

https://www.npr.org/2023/08/26/1195695051/driverless-cars-san-francisco-waymo-cruise

Researchers and practitioners need to develop new techniques for AI forensics.

Develop and validate approaches to acquire and stabilize forensic evidence from AI-enabled cyber-physical systems.

Develop the equivalent of “file hashing” for AI models:

- Detect identical or equivalent models
- Quantify model divergence and impact of model changes

Distinguish model changes and behaviors that are:

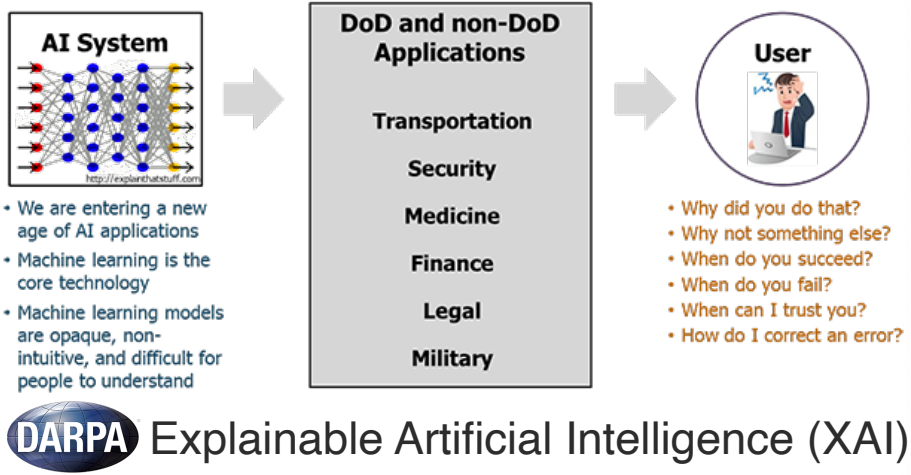
- benign vs. malicious
- emergent vs. directed

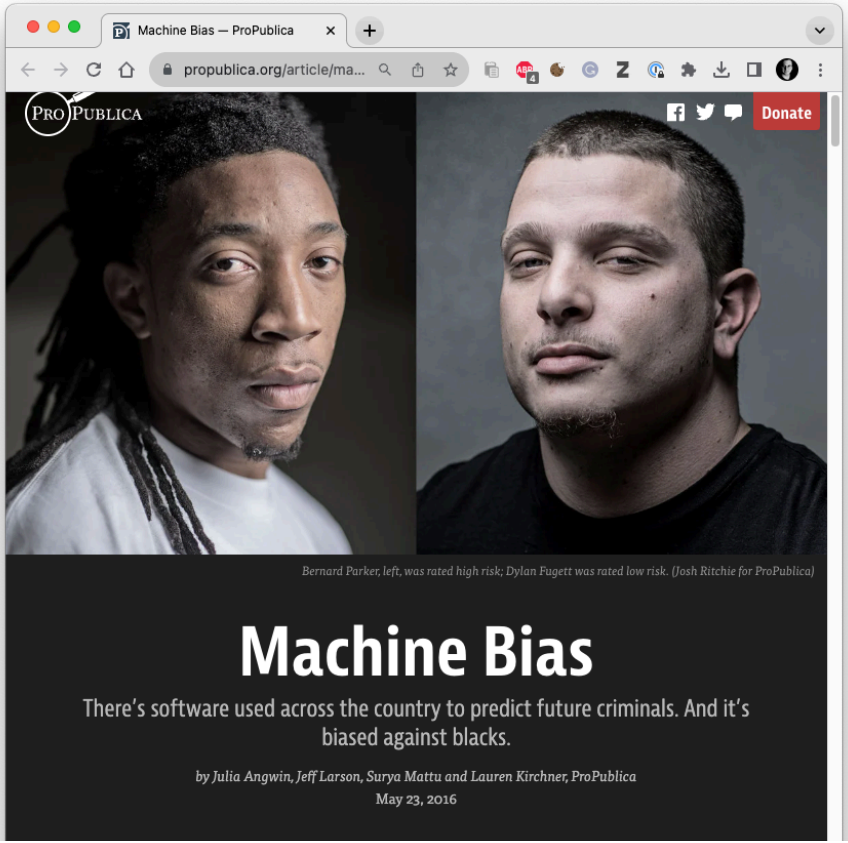
https://en.wikipedia.org/wiki/Slaughterbots

AI forensics is not limited to cyber-physical systems — and it's about more than simply auditing algorithms for fairness or explanation.

There is growing interest in auditing AI systems for "fairness."

Overlaps with Explainable AI (XAI)





ProPublica 2016

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

IBM includes AI detection in AI forensics

IBM Research Focus areas Publications Collaborate Careers Events

24 Jul 2023 Explainer 5 minute read

Did an AI write that? If so, which one? Introducing the new field of AI forensics

IBM researchers are developing AI-text detection and attribution tools to make generative AI more transparent and trustworthy.

24 July 2023
https://research.ibm.com/blog/AI-forensics-attribution

The first targets for "forensics of AI systems" might be AI systems that are used for conventional forensics.

NIST Study Shows Face Recognition Experts Perform Better With AI as Partner

Multidisciplinary study provides scientific underpinnings for accuracy of forensic facial identification.

May 29, 2018

https://www.nist.gov/news-events/news/2018/05/nist-study-shows-face-recognition-experts-perform-better-ai-partner

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

A comparative analysis of human and AI performance in forensic estimation of physical attributes

Sarah Barrington[1,3] & Hany Farid[1,2,3]

Human errors in criminal investigations have previously led to devastating miscarriages of justice. For example, flaws in forensic identification based on physical or photographic evidence are notoriously unreliable. The criminal justice system has, therefore, started to turn to artificial intelligence (AI) to improve the reliability and fairness of forensic identification. So as not to repeat history, it is critical to evaluate the appropriateness of deploying these new AI forensic tools. We assess the feasibility of measuring basic physical attributes in a photo using a state-of-the-art AI system, and compare performance with human experts and non-experts. Our results raise concerns as to the use of current

"Our results raise concerns as to the use of current AI-based forensic identification."

Barrington, S., Farid, H. A comparative analysis of human and AI performance in forensic estimation of physical attributes. *Sci Rep* **13**, 4784 (2023). https://doi.org/10.1038/s41598-023-31821-3

AI and Digital Forensics: An agenda

AI for Digital Forensics

Opportunities:

Make investigators more productive.

- Tool recommendations
- Identify important artifacts
- Correlate information
- Write reports

Needed:

Policies for AI usage and scope.

Validation for bias & error

Digital Forensics for AI

Opportunities:

Understand how AI works and fails

- Pre-deployment certification
- Accountability

Needed:

Improved understanding of:

- How & why deep learning systems work
- How hardware & software interact