

AI and Digital Forensics

**Dr Simson Garfinkel, Chief Scientist
BasisTech**

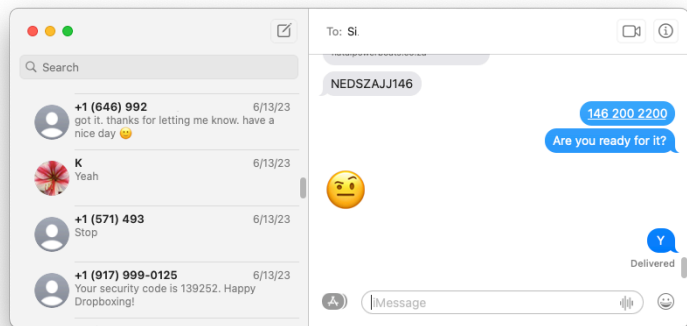
simsong@basistech.com

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

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



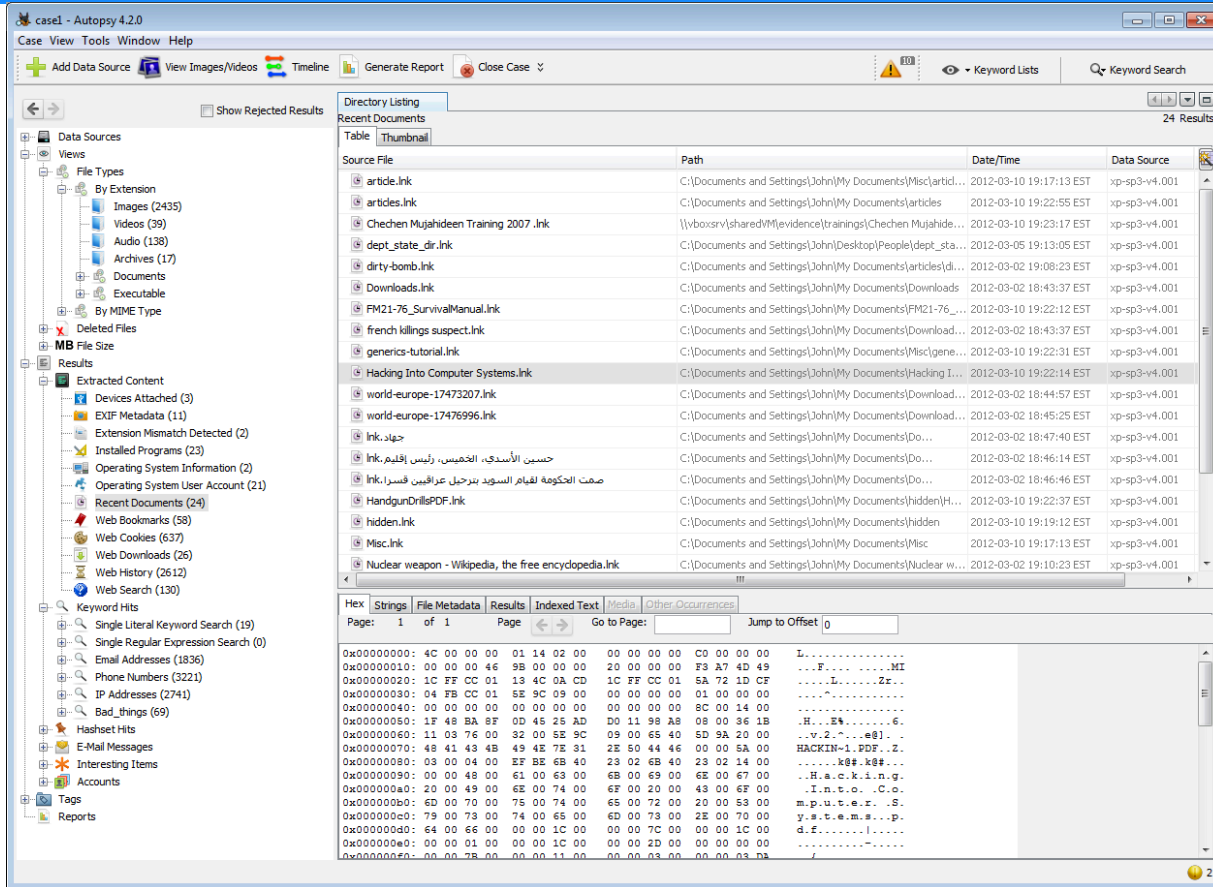
AI for Digital Forensics

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...



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

case1 - Autopsy 4.2.0

Case View Tools Window Help

Add Data Source View Images/Videos Timeline Generate Report Close Case

Show Rejected Results

Directory Listing

Recent Documents

Table Thumbnail

Source File Path Date/Time Data Source

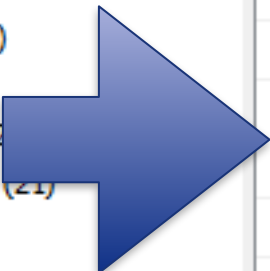
article.link	C:\Documents and Settings\John\My Documents\Misc\articl...	2012-03-10 19:17:13 EST	xp-sp3-v4.001
articles.link	C:\Documents and Settings\John\My Documents\articles	2012-03-10 19:22:55 EST	xp-sp3-v4.001
Chechen Mujahideen Training 2007 .link	\\vboxsrv\shared\VM\Evidence\trainings\Chechen Mujahide...	2012-03-10 19:23:17 EST	xp-sp3-v4.001
dept_state_dir.link	C:\Documents and Settings\John\Desktop\People\dept_sta...	2012-03-05 19:13:05 EST	xp-sp3-v4.001
dirty-bomb.link	C:\Documents and Settings\John\My Documents\articles\di...	2012-03-02 19:08:23 EST	xp-sp3-v4.001
Downloads.link	C:\Documents and Settings\John\My Documents\Downloads	2012-03-02 18:43:37 EST	xp-sp3-v4.001
FM21-76_SurvivalManual.link	C:\Documents and Settings\John\My Documents\FM21-76_...	2012-03-10 19:22:12 EST	xp-sp3-v4.001
french killings suspect.link	C:\Documents and Settings\John\My Documents\Download...	2012-03-02 18:43:37 EST	xp-sp3-v4.001
generics-tutorial.link	C:\Documents and Settings\John\My Documents\Misc\gene...	2012-03-10 19:22:31 EST	xp-sp3-v4.001
Hacking Into Computer Systems.link	C:\Documents and Settings\John\My Documents\Hacking I...	2012-03-10 19:22:14 EST	xp-sp3-v4.001
world-europe-17473207.link	C:\Documents and Settings\John\My Documents\Download...	2012-03-02 18:44:57 EST	xp-sp3-v4.001
world-europe-17476996.link	C:\Documents and Settings\John\My Documents\Download...	2012-03-02 18:45:25 EST	xp-sp3-v4.001
لجاء: جهان	C:\Documents and Settings\John\My Documents\Do...	2012-03-02 18:47:40 EST	xp-sp3-v4.001
حسين الأسدی، الختمیس، رئیس اقلیم: جهان	C:\Documents and Settings\John\My Documents\Do...	2012-03-02 18:46:14 EST	xp-sp3-v4.001
صفه الحكومه للقيام بالسويد بنزول عراقيين قسرا: لجا	C:\Documents and Settings\John\My Documents\Do...	2012-03-02 18:46:46 EST	xp-sp3-v4.001
HandgunDrillsPDF.link	C:\Documents and Settings\John\My Documents\hidden\H...	2012-03-10 19:22:37 EST	xp-sp3-v4.001
hidden.link	C:\Documents and Settings\John\My Documents\hidden	2012-03-10 19:19:12 EST	xp-sp3-v4.001
Misc.link	C:\Documents and Settings\John\My Documents\Misc	2012-03-10 19:17:13 EST	xp-sp3-v4.001
Nuclear weapon - Wikipedia, the free encyclopedia.link	C:\Documents and Settings\John\My Documents\Nuclear w...	2012-03-02 19:10:23 EST	xp-sp3-v4.001

Hex Strings File Metadata Results Indexed Text Media Other Occurrences

Page: 1 of 1 Page Go to Page: Jump to Offset: 0

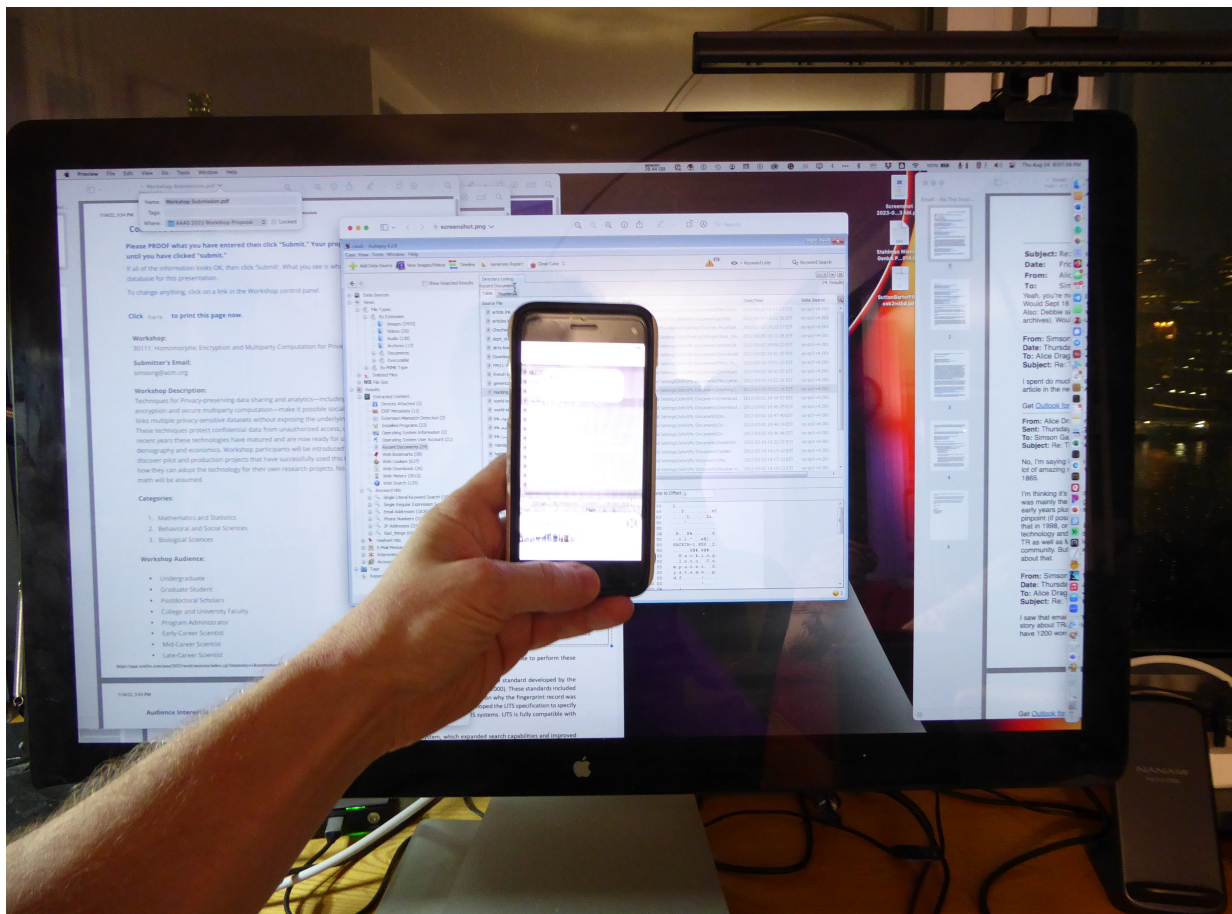
0x00000000: 4C 00 00 00 01 14 02 00 00 00 00 C0 00 00 00 L.....MI
0x00000001: 00 00 00 46 9B 00 00 00 20 00 00 00 F3 A7 4D 49F.....
0x00000002: 1C FF CC 01 13 4C 0A CD 1C FF CC 01 5A 72 1D CFL.....Zr..
0x00000003: 04 FB CC 01 5E 9C 09 00 00 00 00 00 00 01 00 00 00
0x00000004: 00 00 00 00 00 00 00 00 8C 00 14 00 8C 00 14 00
0x00000005: 1F 48 BA 8F 0D 45 25 AD 0D 11 98 A8 09 00 36 1BH.....6..
0x00000006: 11 03 76 00 32 00 5E 9C 09 00 65 40 5D 9A 20 00v.2.....e!..
0x00000007: 48 41 43 4B 49 4E 7E 31 2E 50 44 46 00 00 5A 00HACKIN-1.PDF..Z..
0x00000008: 03 00 04 00 EF BE 6B 40 23 02 6B 40 23 02 14 00k#.#k#..
0x00000009: 00 00 48 00 61 00 63 00 6B 00 69 00 6E 00 67 00H.a.c.k.i.n.g..
0x0000000a: 20 00 49 00 6E 00 74 00 6F 00 20 00 43 00 6F 00I.n.t.o..C.o..
0x0000000b: 6D 00 70 00 75 00 74 00 65 00 72 00 20 00 53 00m.p.u.t.e.r..S..
0x0000000c: 79 00 73 00 74 00 65 00 6D 00 78 00 2E 00 70 00y.s.t.e.m.s...p..
0x0000000d: 64 00 6E 00 00 00 1C 00 00 00 7C 00 00 00 1C 00d.f.....
0x0000000e: 00 00 01 00 00 00 1C 00 00 00 2D 00 00 00 00 00
0x0000000f: 00 00 2B 00 00 00 11 00 00 00 03 00 00 00 03 00/.....

What do these filenames mean in English?



FM21-76_SurvivalManual.Ink	C:\Documents and Settings\Joh
french killings suspect.Ink	C:\Documents and Settings\Joh
generics-tutorial.Ink	C:\Documents and Settings\Joh
Hacking Into Computer Systems.Ink	C:\Documents and Settings\Joh
world-europe-17473207.Ink	C:\Documents and Settings\Joh
world-europe-17476996.Ink	C:\Documents and Settings\Joh
Ink. جهاد	C:\Documents and Settings\Joh
Ink. حسين الأسدي، الخميس، رئيس إقليم	C:\Documents and Settings\Joh
Ink. صمت الحكومة لقيام السويد بترحيل عراقيين قسرا	C:\Documents and Settings\Joh
HandgunDrillsPDF.Ink	C:\Documents and Settings\Joh
hidden.Ink	C:\Documents and Settings\Joh
Misc.Ink	C:\Documents and Settings\Joh

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

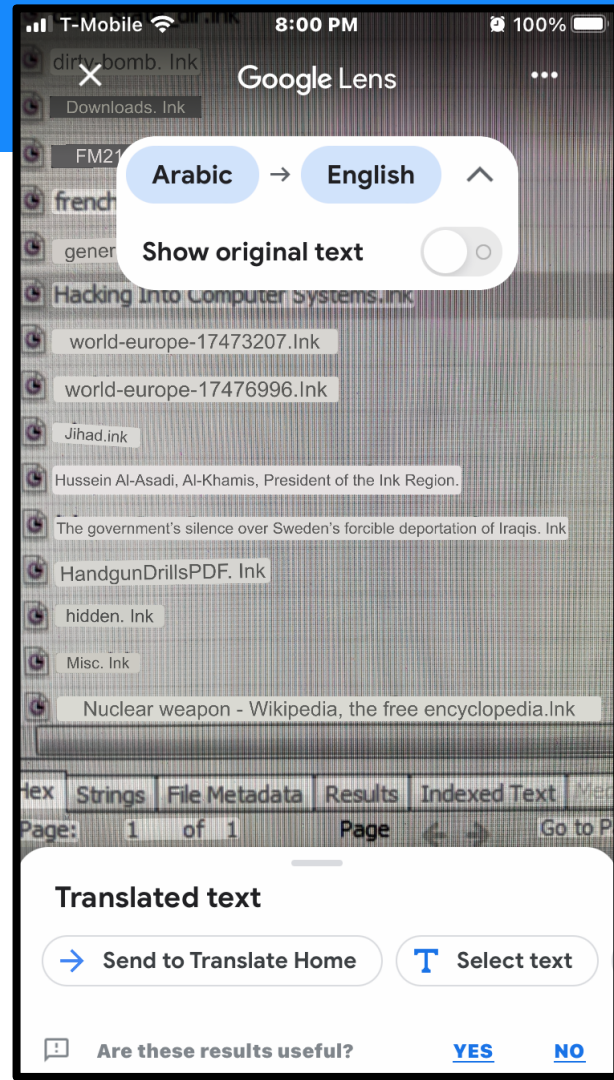


This works surprisingly well... but it's awkward! (and hard to take into court)

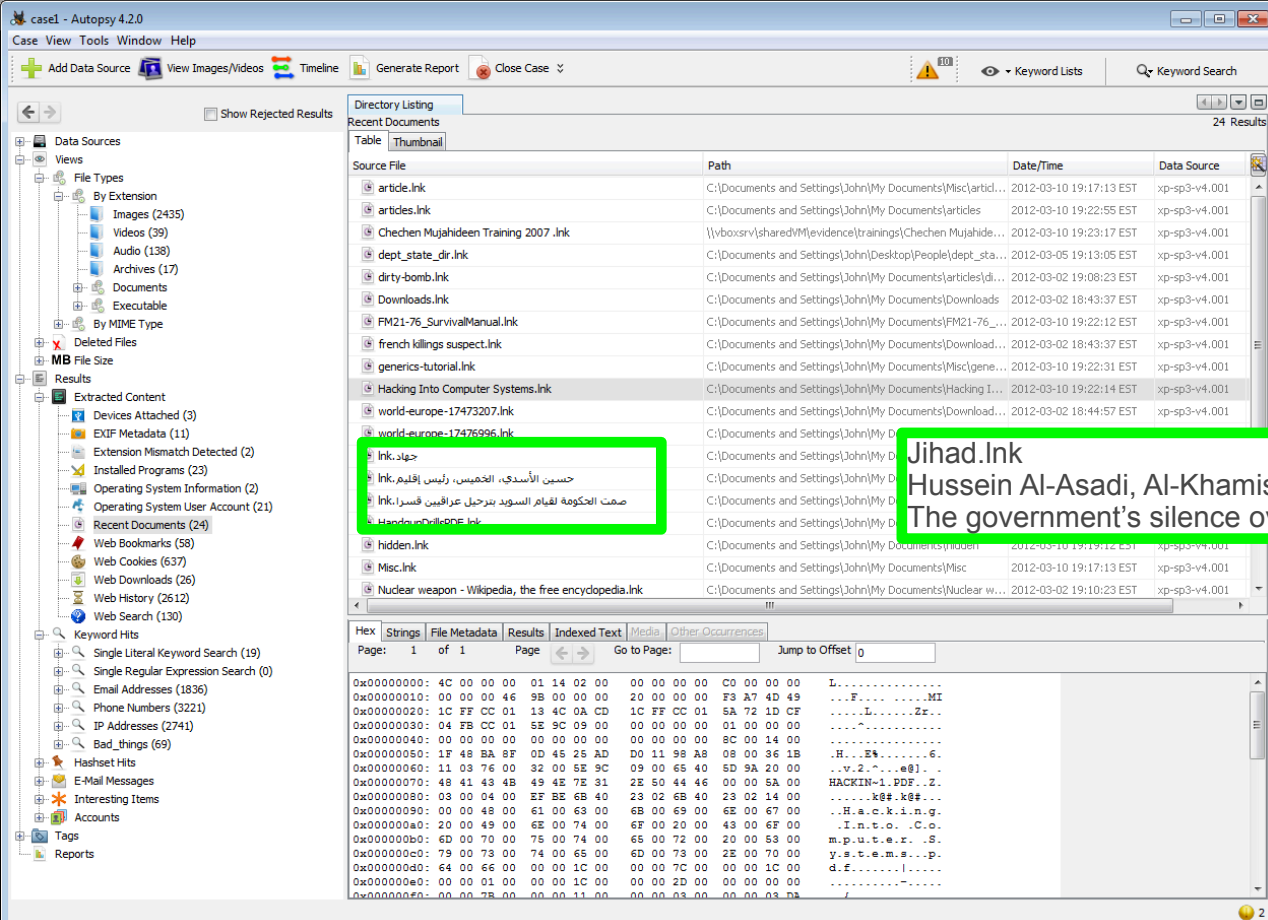
french killings suspect.Ink	C:\Documents and Settings\Joh
generics-tutorial.Ink	C:\Documents and Settings\Joh
Hacking Into Computer Systems.Ink	C:\Documents and Settings\Joh
world-europe-17473207.Ink	C:\Documents and Settings\Joh
world-europe-17476996.Ink	C:\Documents and Settings\Joh
Ink.جهاد	C:\Documents and Settings\Joh
Ink.حسين الأسدي، الخميس، رئيس إقليم	C:\Documents and Settings\Joh
Ink.صمت الحكومة لقيام السويد بترحيل عراقيين قسرا	C:\Documents and Settings\Joh
HandgunDrillsPDF.Ink	C:\Documents and Settings\Joh
hidden.Ink	C:\Documents and Settings\Joh

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.Ink

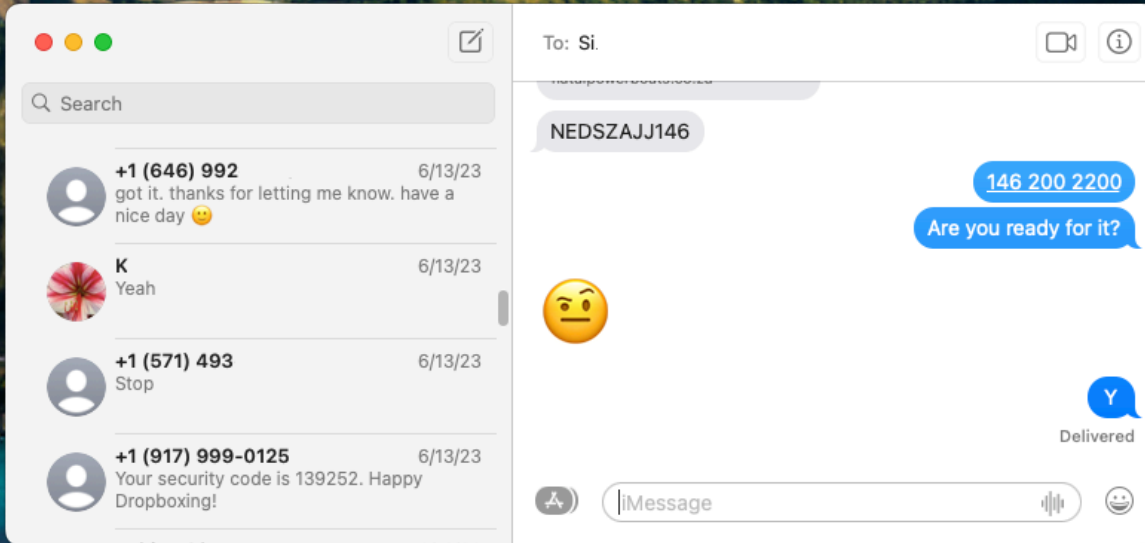
Hussein Al-Asadi, Al-Khamis, President of the Region.Ink

The government's silence over Sweden's forcible deportation of Iraqis.

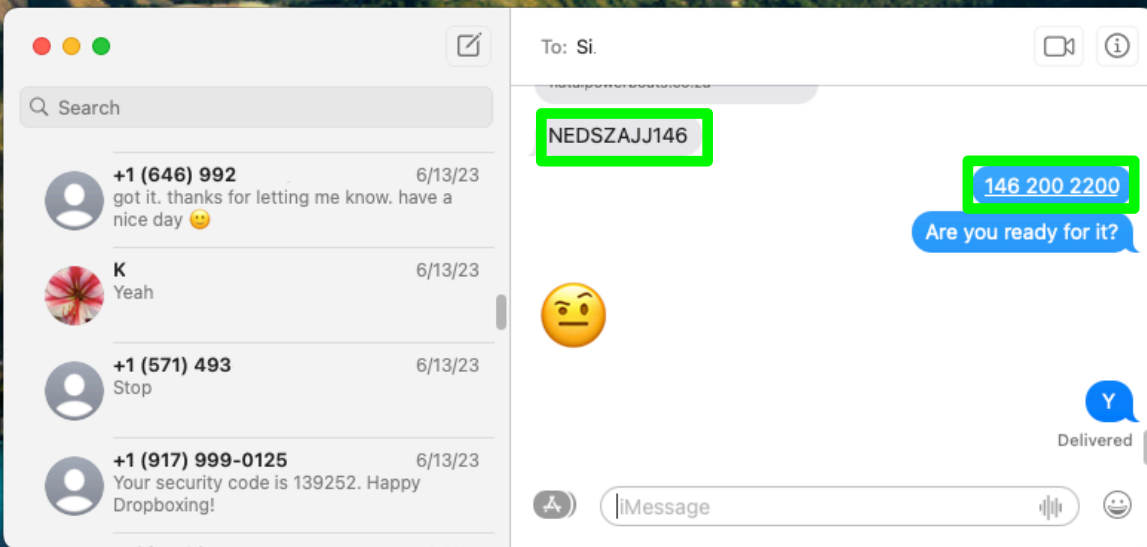
The Forensic CoPilot could provide additional context. Consider a hypothetical case involving Apple Messenger...



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.

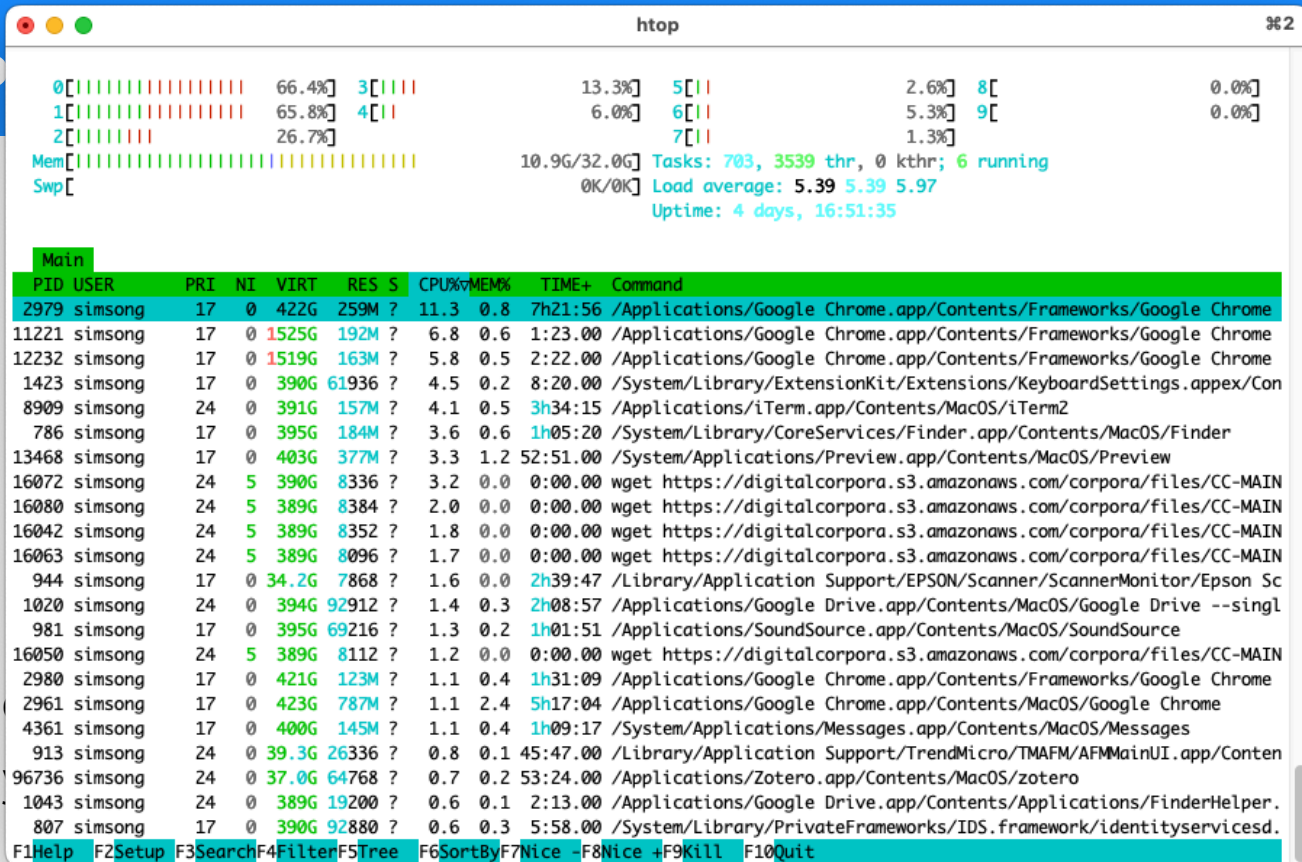
me: system sluggish.

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

me: S

AI: try: **htop**

AI can reco



me: syst

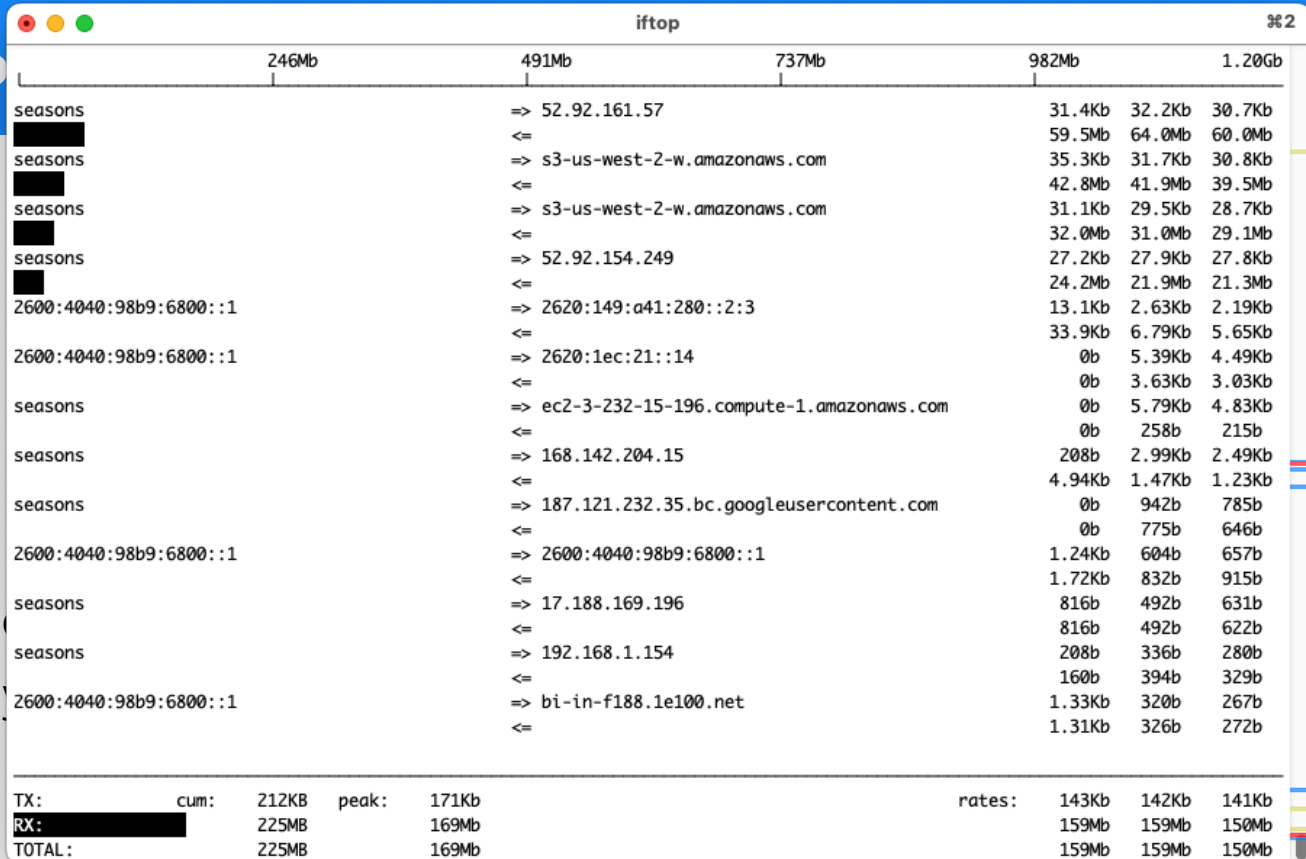
AI: Are

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 ...

SH?

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

WHOIS Details

[Download](#)

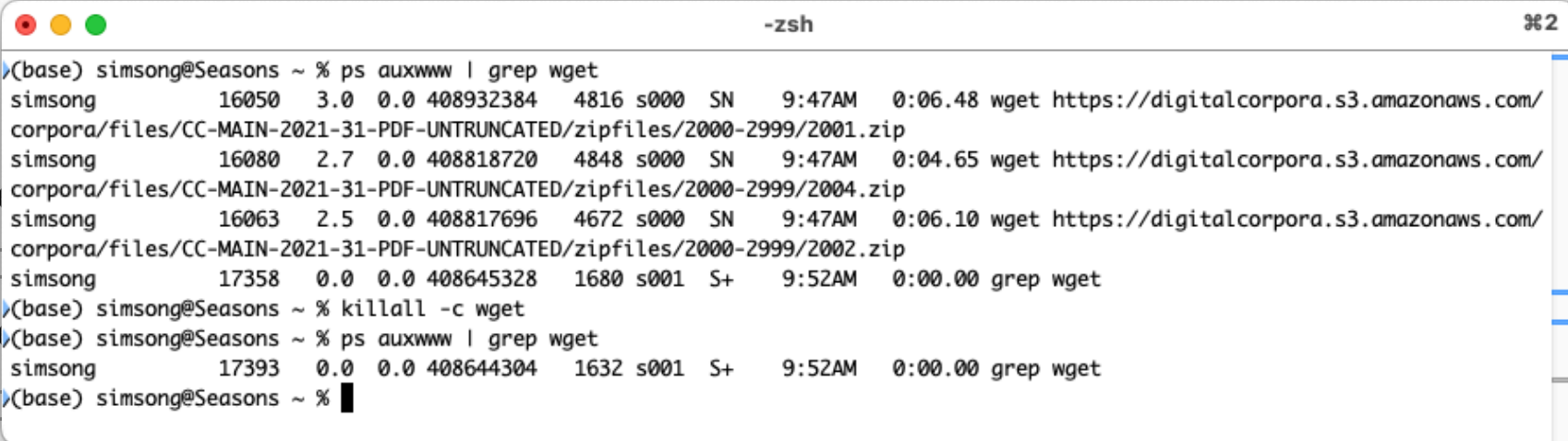
WHOIS Lookup (52.92.161.57)

```
#
# ARIN WHOIS data and services are subject to the Terms of Use
# available at: https://www.arin.net/resources/registry/whois/tou/
#
# If you see inaccuracies in the results, please report at
# https://www.arin.net/resources/registry/whois/inaccuracy_reporting/
#
# Copyright 1997-2023, American Registry for Internet Numbers, Ltd.
#
```

```
NetRange:      52.84.0.0 - 52.95.255.255
CIDR:          52.88.0.0/13, 52.84.0.0/14
NetName:       AT-88-Z
NetHandle:     NET-52-84-0-0-1
Parent:        NET52 (NET-52-0-0-0-0)
NetType:       Direct Allocation
OriginAS:      AS16509, AS14618
Organization:  Amazon Technologies Inc. (AT-88-Z)
RegDate:       1991-12-19
Updated:       2022-03-21
Ref:           https://rdap.arin.net/registry/ip/52.84.0.0
```

```
OrgName:       Amazon Technologies Inc.
OrgId:         AT-88-Z
Address:       410 Terry Ave N.
City:          Seattle
StateProv:     WA
PostalCode:    98109
Country:       US
RegDate:       2011-12-08
Updated:       2022-09-30
```

AI can recommend tools for cyber investigations.



```
-zsh %2
(base) simsong@Seasons ~ % ps auxwww | grep wget
simsong      16050   3.0  0.0 408932384   4816 s000  SN    9:47AM   0:06.48 wget https://digitalcorpora.s3.amazonaws.com/
corpora/files/CC-MAIN-2021-31-PDF-UNTRUNCATED/zipfiles/2000-2999/2001.zip
simsong      16080   2.7  0.0 408818720   4848 s000  SN    9:47AM   0:04.65 wget https://digitalcorpora.s3.amazonaws.com/
corpora/files/CC-MAIN-2021-31-PDF-UNTRUNCATED/zipfiles/2000-2999/2004.zip
simsong      16063   2.5  0.0 408817696   4672 s000  SN    9:47AM   0:06.10 wget https://digitalcorpora.s3.amazonaws.com/
corpora/files/CC-MAIN-2021-31-PDF-UNTRUNCATED/zipfiles/2000-2999/2002.zip
simsong      17358   0.0  0.0 408645328   1680 s001  S+    9:52AM   0:00.00 grep wget
(base) simsong@Seasons ~ % killall -c wget
(base) simsong@Seasons ~ % ps auxwww | grep wget
simsong      17393   0.0  0.0 408644304   1632 s001  S+    9:52AM   0:00.00 grep wget
(base) simsong@Seasons ~ %
```

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.

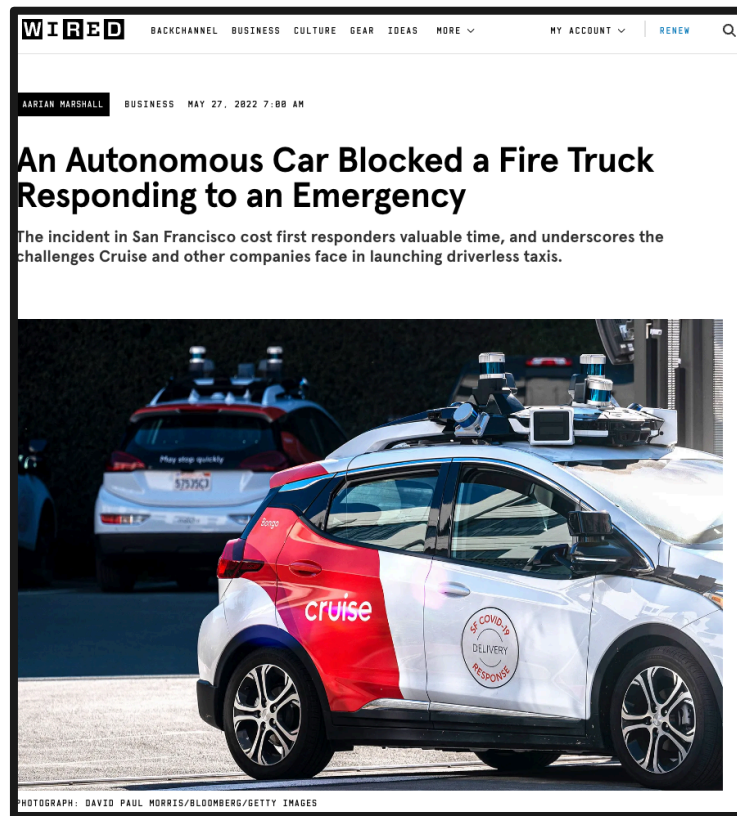
Why did the AI fail?

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

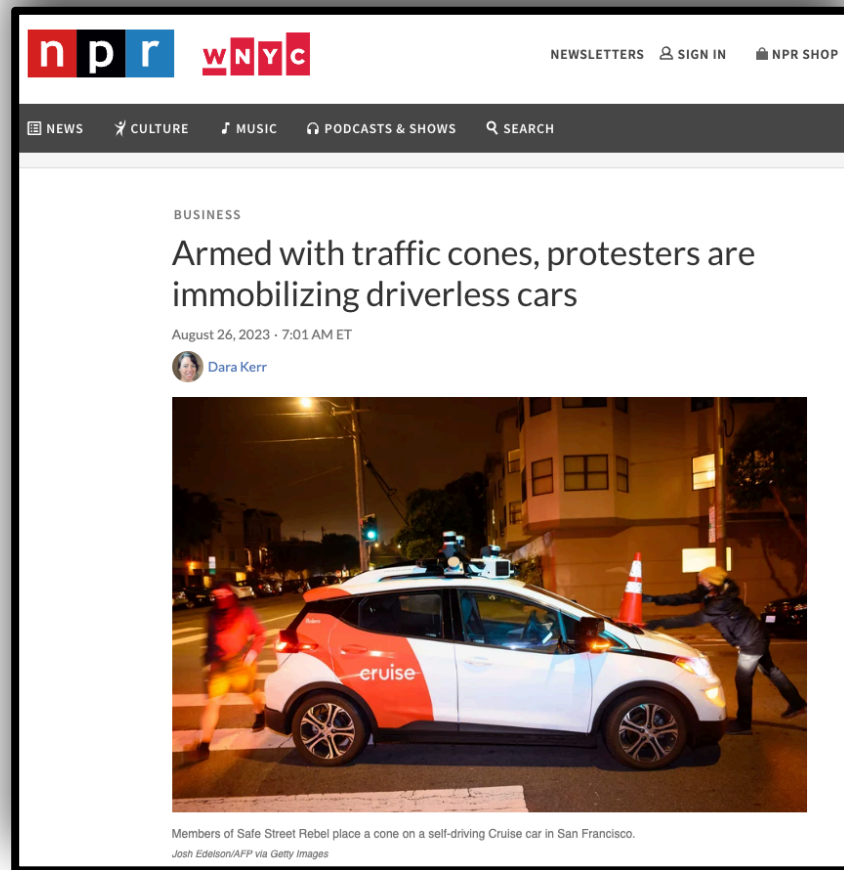
What was the source of this particular incident?

- 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?



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



<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.

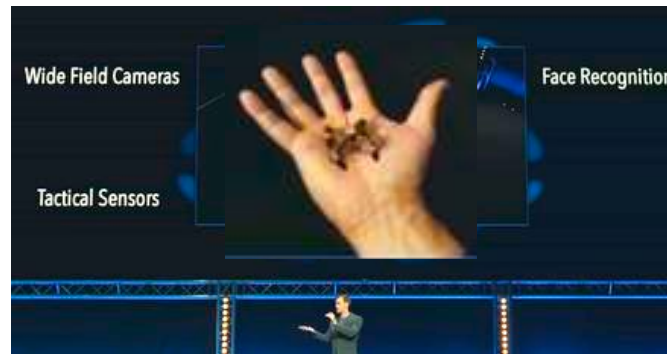
Develop and validate approaches to acquire and stabilize forensic evidence from AI-enabled cyber-physical systems.

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

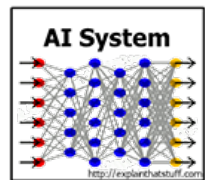


<https://en.wikipedia.org/wiki/Slaughterbots>

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)



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

DoD and non-DoD Applications

Transportation

Security

Medicine

Finance

Legal

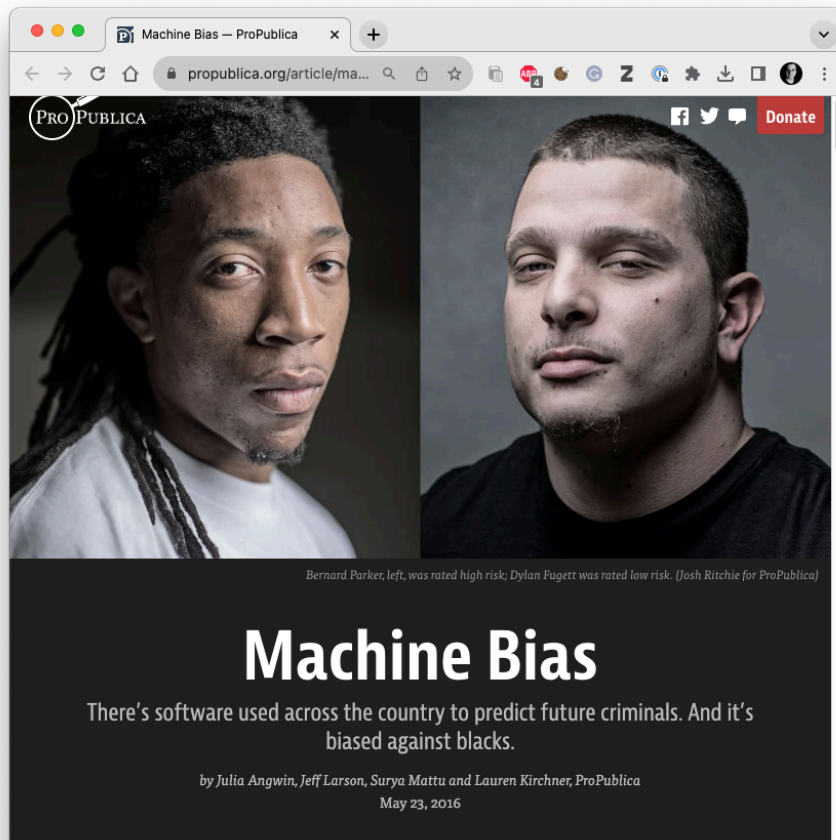
Military




- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?






Explainable Artificial Intelligence (XAI)




IBM includes AI detection in AI forensics

 **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.



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