artificial intelligence in espionage

Carla Zoe Cremer, DPhil, Oxford, Depart. Psychology, Human Information Processing Lab, funded by Future of Humanity Institute

arguments I make here must, if solid in their abstract version, be specified, tailored and applied to a department, country, law enforcement entity or business to have any use or take effect

DAAI: Decisive, Autonomous Artificial Intelligence

AI capabilities that

A) significantly increase national and/or global risks as a result of the degree of **autonomous** (including but not necessitating agentic) code executions, and which are
B) strategically relevant for a nation's defence and/or economic and ideological supremacy

Threat models

- Operational failures (Simonite, 2007)

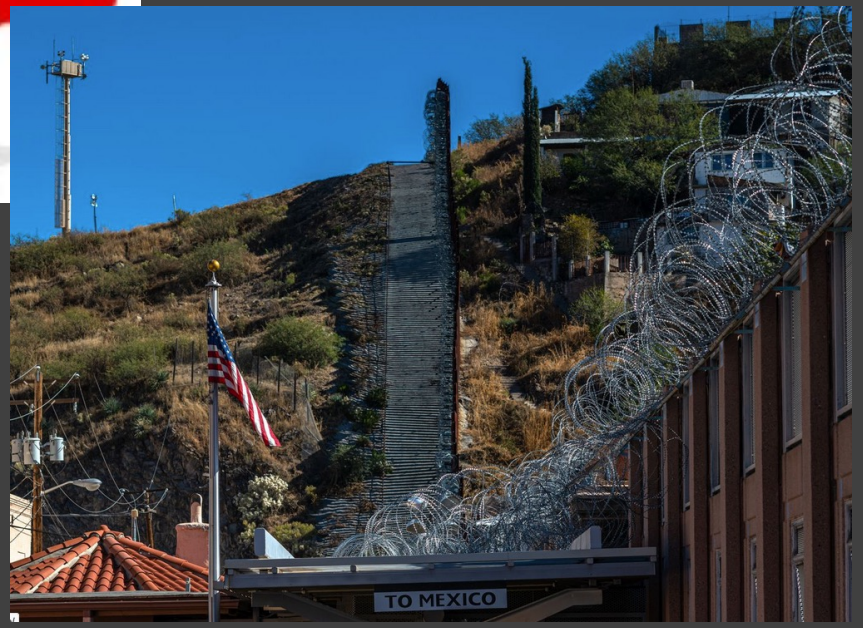
- Institutional failures
  - calcification (Cremer, Oxford AI society May 2022)
  - Lage der Nation 2022

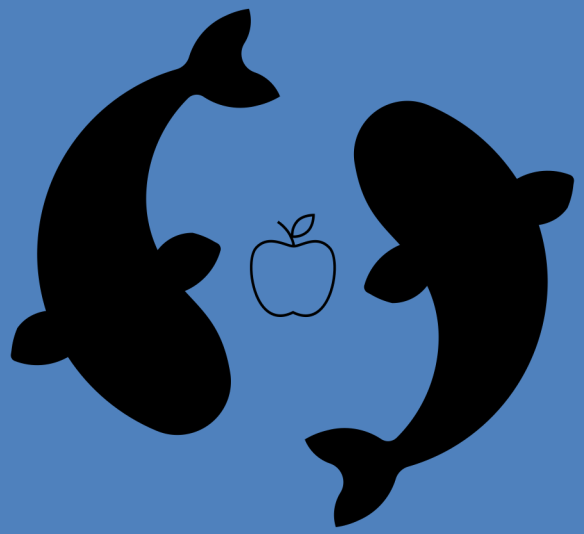
Threat models

- Loss of control (Christian, 2020)

WannaCry 2017

- Amplification of control (Parrish, 2019)





(independent) sources of AI mega projects [incentive + resources]

LLMs as search and compression
- where do the constrains on hallucinations and for counterfactuals come from?(C & Carter, 2023)

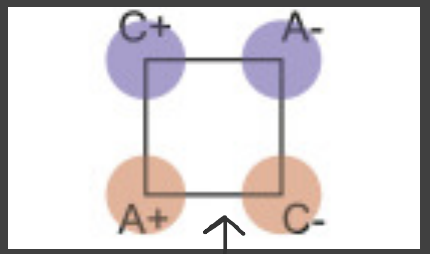
Knowledge Lake + LLM

1. Imperative & Incentive
2. Stakes
3. Funds
4. Data
5. Increasing relevance
6. Speed

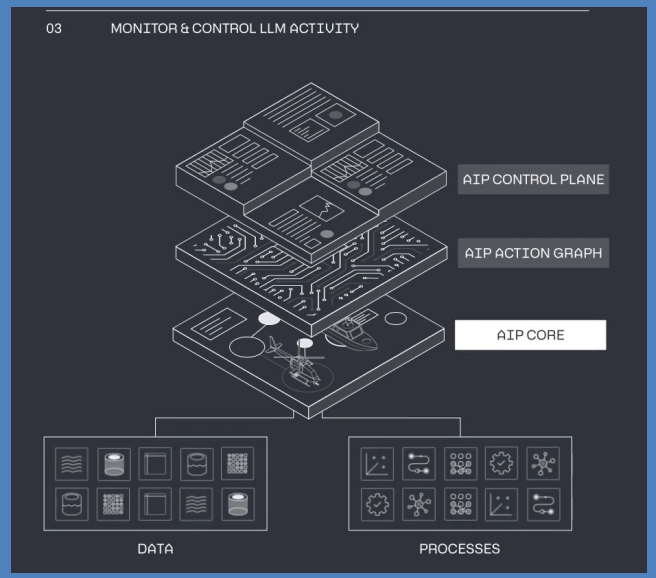
DAAI: Decisive, Autonomous Artificial Intelligence | IC : Intelligence Community

- Imperative, Incentive, Stakes

- information volume (Zegart, 2022), (Fabre 2022)
- active disordering (Joint CSI, 2023)
- information work = representational utility + content



Bernardi, 2020



Palantir

- scientific research and engineering (e.g. reconnaissance satellites)
- cyber defence , e.g. reverse engineering / forensics/ attack detection, anomaly detection
- content verification (e.g. generative vs real images), source verification
- misleading content creation
- dating, geolocating, data-lineage analysis
- translation, sentiment analysis
- summarisations, visualisations of information and communication networks
- formatting, organising, searching multimodal datasets
- integration of different sources, link-analyses, de-anonymisation, tiered trustworthiness
- social engineering, persuasion and information extraction
- shortening analysis steps during combat

DAAI: Decisive, Autonomous Artificial Intelligence | IC : Intelligence Community

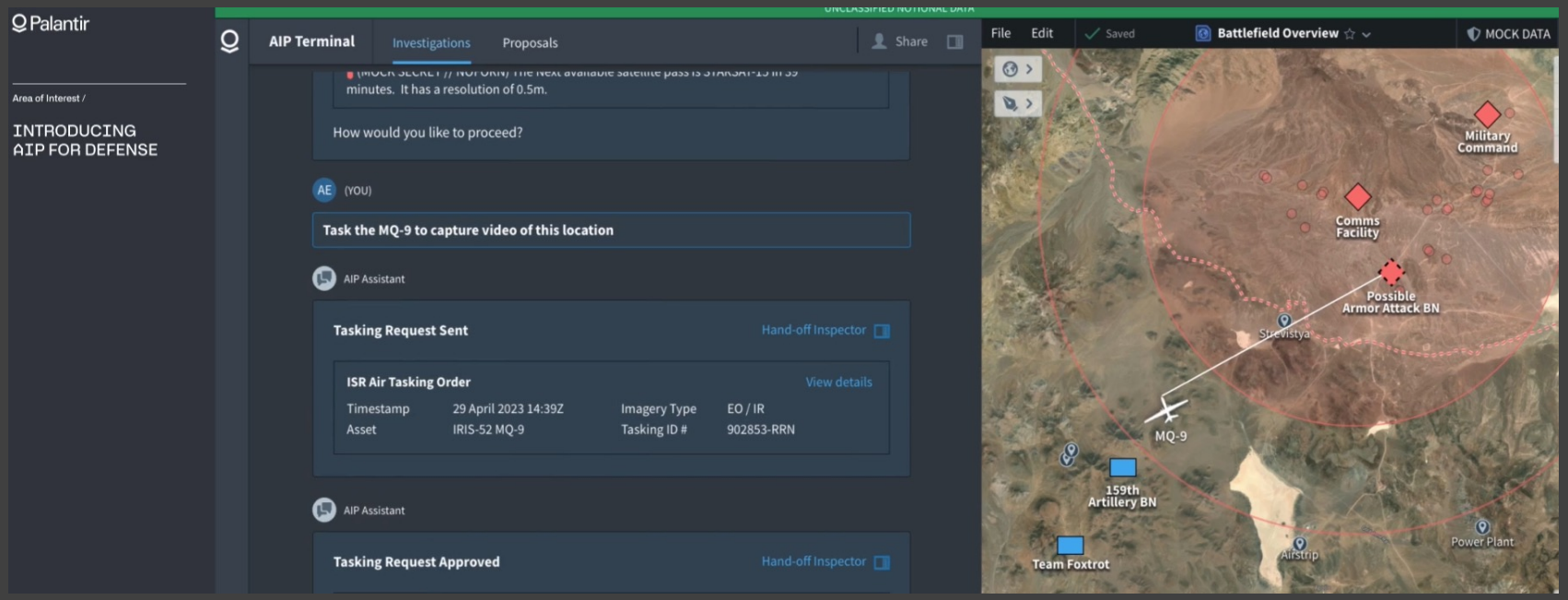
Particular constraints/predictions imposed by machine learning applications (Janjeva et al., 2023)

- data intensive
- multi-modal
- bulk collection (Anderson, 2016)
- increases the pressure to retain data for longer, diversify data and collect more data

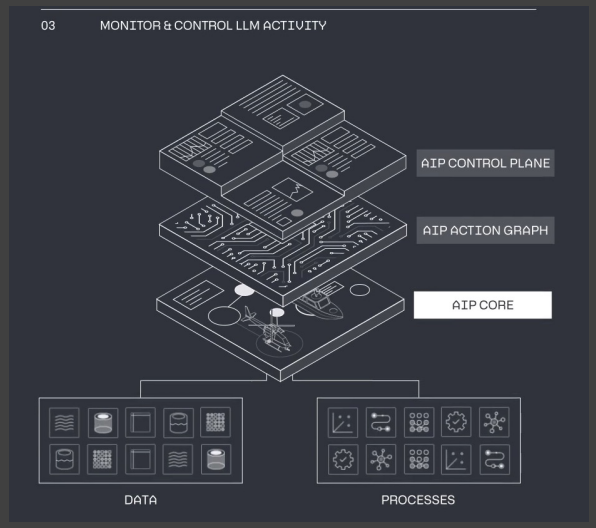
- can legitimise and leak surveillance technology and policy (Turse, 2017), (Parkinson et al., 2019)
- increases risks of
    - data breach and hacks (e.g. Office of Personnel Management Data Breach, 2015 or Snowden 2013), emergency powers extension (Kemp, 2021)

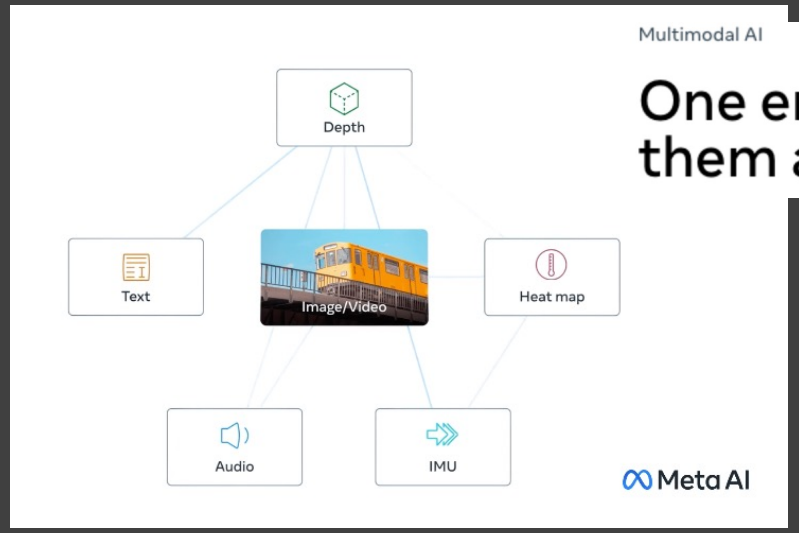
Imperative, Incentive, Stakes

- Oppenheimer moment - (Karp, 2023)
- Weinbaum & Shanahan, 2018
- Long history of waiting for AI and ongoing work (Moran, Burton and Christou 2023)



(*Palantir AIP*,2023)



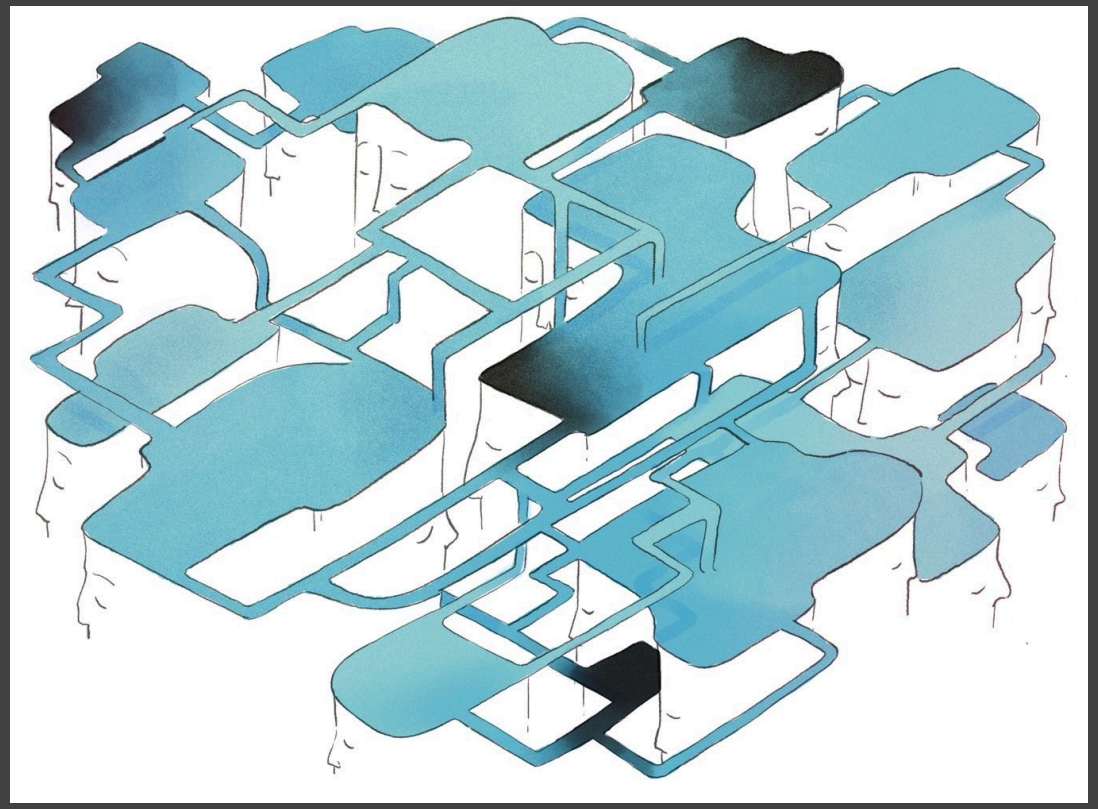


(*ImageBind*, 2023)

US $93.7 billion <

China. (2023). Intelligence and Security Committee of Parliament. https://isc.independent.gov.uk/wp-content/uploads/2023/07/ISC-China.pdf

curate, filter a polluted epistemic landscape (SemaFor, DARPA, Air Force Reserach Lab, DeepMedia)



@magdalenadomeit

Elite epistemic management (Speri, 2023, Seib 2021) &
the changing nature of warfare (Tsamados et al. 2023)

OSINT?

ICs as political actors?

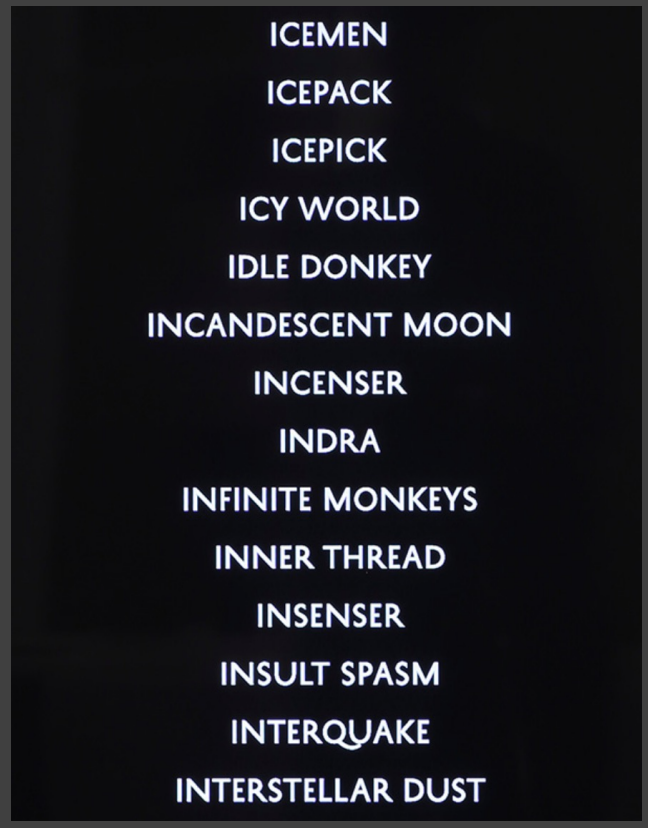
Another way: High stakes contexts to stress test algorithms instead of safety washing?

Risk trade-offs : reducing one by means of DAAI increases others?

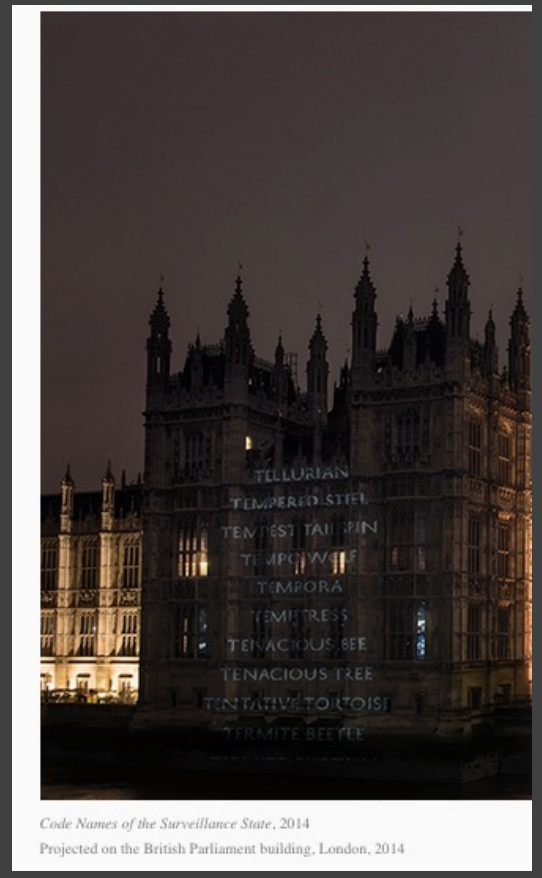
Technologically informed oversight?

More general responsibilities for epistemic infrastructure?

IC and the hope of epistemic coordination?



Trevor Paglen 2014



Code Names of the Surveillance State, 2014
Projected on the British Parliament building, London, 2014

A New Generation of Intelligence: National Security and Surveillance in the Age of AI. (2023, October 16).

Anderson, D. (2016). _Investigatory Powers Bill: Bulk powers review_. GOV.UK.

Army Vantage's CRRT completes rollout to all Army components | PEOEIS. (n.d.). Retrieved November 15, 2023

Artificial Intelligence and UK National Security: Policy Considerations. (2023a, October 16).

Babuta, A., Oswald, M., & Janjeva, A. (n.d.-b). _Artificial Intelligence and UK National Security_.

Boháček, M., & Farid, H. (2022). Protecting world leaders against deep fakes using facial, gestural, and vocal mannerisms. _Proceedings of the National Academy of Sciences_, _119_(48), e2216035119.

C, A., & Carter, R. (2023). _Large Language Models and Intelligence Analysis_ [Expert Analysis]. CETaS.

Christopher Nolan says AI is our Oppenheimer moment. (2023, July 17). Archive.Ph. https://archive.ph/STqtF
German parliament to stop using fax machines. (2021, January 15). _POLITICO_.

Hardy, J. (2016). Targeting thresholds: The impact of intelligence capability on ethical requirements for high-value targeting operations. In J. Galliott & W. Reed (Eds.), _Ethics and the future of spying_ (pp. 177–190). Routledge.

Hernandez, E. (2022, March 30). _What is Operation Lone Star? Gov. Greg Abbott's controversial border mission, explained._ The Texas Tribune.

Roy, D., Cheatham, A., & Klobucista, C. (2023). *How the U.S. Patrols Its Borders*. Council on Foreign Relations.

ImageBind: Holistic AI learning across six modalities. (2023). Meta

Intelligence and Security Committee of Parliament: China. (2023). Intelligence and Security Committee of Parliament.

Intelligence Community Spending Trends. (2023). Congressional Research Service.
https://sgp.fas.org/crs/intel/R44381.pdf

Janjeva, A., Calder, M., & Oswald, M. (2023). _Privacy Intrusion and National Security in the Age of AI_.

Karp, A. C. (2023, July 25). Opinion | Our Oppenheimer Moment: The Creation of A.I. Weapons.

Carla Cremer , 15 11 2023

Karp, A. C. (2023, July 25). Opinion | Our Oppenheimer Moment: The Creation of A.I. Weapons. _The New York Times

Kemp, L. The “Stomp Reflex”: When governments abuse emergency powers

Lin, P., & Ford, S. (2016). _I, spy robot: The ethics of robots in national intelligence activities_.

LLC, E. S. of A. (2019)U.S. Customs and Border Protection, Tohono O’odham Nation Agree On Border Security Solution by Elbit Systems of America_.

Moran, C. R., Burton, J., & Christou, G. (2023a). The US Intelligence Community, Global Security, and AI: From Secret Intelligence to Smart Spying. _Journal of Global Security Studies_, _8_(2), ogad005. [https://doi.org/10.1093/jogss/ogad005]

Moran, C. R., Burton, J., & Christou, G. (2023b). The US Intelligence Community, Global Security, and AI: From Secret Intelligence to Smart Spying. _Journal of Global Security Studies_, _8_(2), ogad005. [https://doi.org/10.1093/jogss/ogad005]

Office of Personnel Management data breach. (2023a). Wikipedia.

Omand, S. D. (2020). _How Spies Think: Ten Lessons in Intelligence_. Viking.

Parrish, W. (2019, August 25). _The U.S. Border Patrol and an Israeli Military Contractor Are Putting a Native American Reservation Under “Persistent Surveillance.”_ The Intercept. [https://theintercept.com/2019/08/25/border-patrol-israel-elbit-surveillance/]

Robust artificial intelligence for active cyber defence. (n.d.). The Alan Turing Institute.

Semantic Forensics. (n.d.). Retrieved November 14, 2023, from https://www.darpa.mil/program/semantic-forensics

Shachtman, N. (2007). Robot Cannon Kills 9, Wounds 14. _Wired_.

Simonite. (2007). _“Robotic rampage” unlikely reason for deaths_. New Scientist. Retrieved November 14, 2023, from [https://www.newscientist.com/article/dn12812-robotic-rampage-unlikely-reason-for-deaths/]
Sky News (Director). (2023, July 19). _MI6 Boss delivers speech on Ukraine, AI technology and Russia_. https://www.youtube.com/watch?v=YFF7cc3jeWM

Surveillance by intelligence services: Fundamental rights safeguards and remedies in the EU - Volume II: field perspectives and legal update. (2017, October 18). European Union Agency for Fundamental Rights. [http://fra.europa.eu/en/publication/2017/surveillance-intelligence-services-fundamental-rights-safeguards-and-remedies-eu](

Tsamados, A., Floridi, L., & Taddeo, M. (2023). The Cybersecurity Crisis of Artificial Intelligence: Unrestrained Adoption and Natural Language-Based Attacks. _SSRN Electronic Journal_. [https://doi.org/10.2139/ssrn.4578165]

Valle, G. D. (2022, April 21). _The Most Surveilled Place in America_. The Verge. [https://www.theverge.com/c/23203881/border-patrol-wall-surveillance-tech]

Waters, B. G. (n.d.). _An International Right to Privacy: Israeli Intelligence Collection In The Occupied Palestinian Territories_. _50_.

Weinbaum, C., & Shanahan, J. N. T. (2018). _Intelligence in a Data-Driven Age_.

Zegart, A. (2022). _Spies, Lies, and Algorithms_.

Carla Cremer , 15 11 2023