CYBERNETICA

On Collaborative Artificial Intelligence and Cybersecurity Operations Between Allies

Research report

Version 1.0

27.10.2023

D-2-539

Public

Authors: Dan Bogdanov, Cybernetica AS

Cybernetica research reports are available online at
https://research.cyber.ee/

Mailing address:
Cybernetica AS
Mäealuse 2/1
12618 Tallinn
Estonia
info@cyber.ee

1 The case for collaborative AI

1.1 Defining collaborative AI tools

Artificial intelligence (AI) systems provide us with predictions that are based on the previous knowledge used in the training of the AI system [1]. Such predictions help us build decision support tools that support humans in making choices. Through the prediction ability of the AI system, these choices are informed by all the knowledge that came before and the human might make a better decision. The effects of such assisted decision making capability have been demonstrated for healthcare, although also with limitations [2].

The advent of large language models has improved the perceived capability of AI systems to perform tasks that can be considered human. Thanks to an enormous amount of training data, consisting of the collective digital knowledge of our societies, new AI systems can synthesise (predict?) text, images and sound with impressive results. The training information has been collected from all available sources, public and private, with repercussions resulting from copyright issues.

Collaboration on AI has the primary benefit of building new tools trained from the knowledge bases of all collaborators. Using data from both sides makes the tools applicable to the contexts of all parties, reducing bias. More data in the inputs increases the chances of having useful properties from the resulting AI models. The shared model could be used and improved with feedback by all collaborators.

1.2 Potential of data collaborations across domains

In cybersecurity, joint situational awareness will help allies react to threats on the Internet, a globally shared cyberspace. Specifically, sharing intelligence on new threat vectors and live attacks in progress could result in strengthening infrastructure and joint investigation and takedown of the attacker. Joint anomaly detection on service logs could open up new understanding of globally coordinated cyberattacks.

In healthcare, AI will be helping the physicians correctly diagnose and treat complex and rare diseases that might otherwise not have experience with. AI could learn from the medical experience of hundreds of thousands of doctors having treated hundreds of millions patients from different age groups, genders and races. A clinical decision support tool (CDST) trained on this knowledge will achieve higher precision on wider population, avoiding biased decisions that would result from skewed training data.

Similar potential exists for e-government, law enforcement and intelligence applications. Whenever we need a wider coverage of data or to bring in global know-how, data collaborations are the way forward.

1.3 Value of allied collaboration

If citizens, companies and the government can collaborate on data applications without being afraid that the data would be used against them, there will be adoption new services, innovation and growth. If we, as allies, can collaborate on joint AI tools and applications, there will be increased trust in each other, leading to closer collaboration, especially if all parties can retain control of their confidential data and proprietary intellectual property.

Being a partner in an allied collaboration like the one envisioned will also enhance the country's reputation and credibility in the field of AI research and development. Increasing the maturity and knowledge of the societies involved in the joint AI work will also benefit the populations. Working together, we can identify and address ethical and social implications of AI. We can also ensure that AI systems are developed in a responsible and transparent manner—as collaborations require transparency and responsibility towards each other.

There are also economic aspects to note. Collaborative AI development means that each participant must contribute skills, data and infrastructure to the effort. For fairness, these contributions should be in alignment with the resources available to the participants. However, the ones participating more would also be entitled to a greater growth in the macroeconomic metrics once the time to reap the rewards of collaboration comes. A faster time-to-market of new high-quality, unbiased AI systems will benefit the whole world.

2 Challenges in development and operations of joint AI systems

2.1 Data regulations and policies

Above, we set the ability to pull together AI training data on many people, organisations or countries as a prerequisite to achieving the expected results. One of the first challenges to tackle will be regulations related to data and artificial intelligence.

Within the European Union, geolocalisation regulations in member states have largely been removed during the development of the Digital Single Market. However, challenges remain in data collaborations with countries outside the EU. The European Commission has made adequacy decisions with regard to the level of data protection for multiple countries [1] Still, some important data transfer routes remain challenging.

Most notably, there have been multiple attempts at setting up a personal data transfer regime between European Union and the United States of America. One of the most recent ones, the Privacy Shield, was deemed illegal by the Court of Justice of the European Union in July 2020[2]. Since then, a Trans-Atlantic Data Privacy Framework has been announced [3] This warranted an open letter from noyb, a European organisation related to mr Max Schrems, who filed the case that brought down Privacy Shield [4] highlighting the potential shortcomings of the new framework.

In artificial intelligence, regulatory activities are underway. The European Union is developing the AI Act [5]. While still under discussion, it is expected to set restrictions on some AI applications, e.g., the real-time use of biometric identification in the public space.

In cybersecurity, the allied collaboration landscape is largely unregulated, both between civilian and defence stakeholders. There is notable room for development in the real-time exchange of cybersecurity threat intelligence that would be immediately actionable. For that, trust and standards need to be established that would help classify data according to the policies of the data holders, especially in the defence setting. Integrating data exchange and AI systems into military infrastructures, while not necessarily much harder than integrating into hospitals, is still a notable challenge. Existing regulations and security standards have not been designed with today's technological controls in mind.

[1] EU data protection adequacy decisions. https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en Last accessed October 12th, 2023.

[2] Judgment of the Court. July 16th, 2020. https://curia.europa.eu/juris/document/document.jsf?docid=228677&doclang=en Last accessed October 12th, 2023.

[3] European Commission and United States Joint Statement on Trans-Atlantic Data Privacy Framework. March 22nd, 2022. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2087 Last accessed October 12th, 2023.

[4] Open Letter on the Future of EU-US Data Transfers. May 23rd, 2022. https://noyb.eu/en/open-letter-future-eu-us-data-transfers Last accessed October 12th, 2023.

[5] EU AI Act: first regulation on artificial intelligence. June 14th. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence Last accessed October 12th, 2023.

2.2 Technological gap

There is a significant opportunity to overcome the problems related to data sharing hesitations by adopting technological advancements in security and privacy technologies. Technological innovations offer a paradigm shift by reducing the dependency on high trust levels and introducing robust control mechanisms. Specifically, technologies such as Secure Multi-Party Computation (MPC), Federated Learning (FL), Confidential Computing/Trusted Execution Environments (TEE), Homomorphic Cryptography (FHE) exhibit properties that support distributed control by allied collaborators.

These technologies have been developed by countries around the world. In the United States, DARPA, IARPA, NIH and NSF, among others, have supported research into the technologies. The European Commission research and development programs (Framework Programme 7, Horizon 2020, Horizon Europe among others) have done the same. Venture capital powered startup companies have been increasing the maturity of the technologies and the public and private sector procurement market for the technologies is forming. There are also national [3] and cross-border [4] initiatives to support the uptake of these technologies.

The advancement and implementation of these technological tools are anticipated to revolutionise collaborative endeavours by enabling the development of joint decision-support models. These models would capitalise on the collective experience of coalition members without necessitating the explicit sharing of sensitive information or expertise. Such tools will be especially beneficial for entities that have encountered difficulties in initiating machine learning projects that cross organisational or national boundaries.

To substantiate the utility of these emergent technologies, imagine them in a healthcare-inspired context. Using the technologies, we could build a Collaborative Clinical Decision Support Tool (CDST) that assists clinicians in making diagnostic decisions, based on an AI model. Implementation of such a tool is expected to enhance the medical decision-making efficacy of healthcare practitioners in remote or field locations. In particular, the CDSTs will enable medical personnel stationed in remote areas to receive healthcare services commensurate with those available in leading medical facilities. Moreover, the collaborative approach to model development will enrich the machine learning algorithms with data from a more diverse population, thereby enhancing their general applicability and minimising biases and inaccuracies.

2.3 Infrastructure gap

AI infrastructures (data centers with neural network accelerators) are built internationally at an increasing pace. They are critical for training large models. However, more niche models can be trained with common infrastructures.

However, the collaborative secure computing technologies described above will need compatible infrastructure controlled by each stakeholder. This will reduce lock-in, dependence on centralised infrastructure and support localised infrastructure.

Today, the most popular cloud service providers are from the US. Platforms like Microsoft's Azure, Amazon Web Services, Google Cloud, IBM Cloud and others are providing various high-level tools, including for securing computing tools. However, none of them currently provide a scalable cross-cloud secure computing platform, much like the one envisioned in [5, 6].

3 A path forward

3.1 Lighthouse projects will show the way

Seeing is believing, so we should conduct trials and pilot projects that demonstrate the new capabilities in a real-world context. For this, we need the following:

1. stakeholders willing to commit resources to the project,
2. technologies and standards that enable to project and
3. a project idea is perceived as inspiring yet not too risky.

Below, we discuss two lighthouse project concepts that could develop new capabilities in artificial intelligence or cybersecurity.

3.2 Allied cybersecurity pilot concept: threat intelligence exchange network

What if organisations in allied nations could let each other know about ongoing cyberattack campaigns in a systematic and machine readable way? This could allow for swifter response, clearer attribution or even semi-automated counter-operations (whether defensive or offensive).

This capability requires two main components. First, a standardised messaging scheme will ease communication and allow for the interoperability of tools deployed by each ally. Second, a new breed of secure data collaboration tools are needed that provide provable guarantees that the data cannot be processed in a way not agreed to by their owners.

Steps toward the former have been taken in the VORMSI project conducted by the United States Air Force Research Laboratory (USAFRL) and Cybernetica, under the supervision of the Estonian Ministry of Defence and the US Department of Defense[1]. Similarly, European nations are working towards similar goals in the ECYSAP project[2] And the Estonian-South Moravian cybersecurity collaboration project CHESS is an example of a clear desire by two states to share experience and skills[3].

Enabling technologies for this capability have been developed in the NAPLES project[4] of the DARPA Brandeis program[5] which, in turn, built on work done in the SEVILLA project[6] of the DARPA PROCEED[7].

The end goal is to have the capability running not just between Estonia and the US or Estonia and other European countries, but rather the whole of NATO.

[1] VORMSI project brief. Cybernetica website. https://cyber.ee/research/projects/vormsi Links on this page last accessed on October 19th, 2023.

[2] European Cyber Situational Awareness Platform. https://www.ecysap.eu.

[3] Cyber-security Excellence Hub in Estonia and South Moravia (CHESS). https://cyber.ee/research/projects/chess.

[4] NAPLES project brief. Cybernetica website. https://cyber.ee/research/projects/naples.

[5] DARPA Brandeis. https://www.darpa.mil/program/brandeis.

[6] SEVILLA project brief. https://cyber.ee/research/projects/sevilla.

[7] PROgramming Computation on EncryptEd Data (PROCEED). https://www.darpa.mil/program/programming-computation-on-encrypted-data.

3.3 Allied AI pilot concept: joint health statistics between Europe and US using technical privacy enforcement

What if we could pool knowledge between European and North American health researchers into models? We could learn to better understand conditions that reduce he number of years lived healthily and create economic losses in Western countries (neurodegenerative diseases, obesity, cancer and others). We could build clinical decision support tools that bring the world's best knowledge into the hands of every physician.

The task in the title might seems simple, but the complexity is in the details. First - Europe and United States have been working on regulating personal data transfer for a while and still, there are significant challenges. If we could get data transfers to work from EU to the US, the other directions may be easier. Also, this direction could be quite the economic boon on both sides as European stakeholders could stop worrying about the use of US infrastructure (e.g., cloud) for such projects.

Given that health data is a special category in the European General Data Protection Regulation, we could start with simple study that combines aggregated data from reasonably large populations on both sides. This way, the first study uses low-risk data, for which we can quite easily show, that they are not directly linkable to a single person on either side.

Even with such low-risk data, the pilot project would reach a greater impact in the long term by experimenting with privacy enhancing technologies. If these technologies receive scrutiny during this pilot, the involved parties may understand that they have the potential to also protect more sensitive records better than the technologies in use today. The following pilots could then already reach for higher targets.

The UK-US Privacy Enhancing Technology challenge took good steps, but was restricted to organisations in the UK and US. Since Brexit, we would need a new EU-US effort on this.

Cybernetica's PAI-MACHINE project funded by the Office of Naval Research Global is working towards building allied AI systems based on machine-optimised cryptographic protocols[8]. In Europe, the TEADAL project is tackling the challenge from an energy and deployment efficiency direction[9].

[8] Synthesis of machine-optimised cryptographic protocols with applications in secure machine learning systems (PAI-MACHINE). http:/https://cyber.ee/research/projects/pai-machine.

[9] Trustworthy, Energy-Aware federated DAta Lakes along the computing continuum (TEADAL). https://www.teadal.eu.

4 Conclusion

Deep trust and collaboration in the digital world benefits from treatment as equals and sovereignty over their digital identities and data. Not all allied nations on Earth have the same resources available, so sharing is critical. New security technologies increase the control collaborators have over their data and identity information. We encourage stakeholders to reach out to their partners in other nations to initiate secure data sharing and cybersecurity projects to strengthen relations.

5 References

[1] Ajay Agrawal, Joshua Gans, and Avi Goldfarb. *Power and Prediction: The Disruptive Economics of Artificial Intelligence*. Harvard Business Review Press, 2022.

[2] Nikhil Agarwal et al. *Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology*. Tech. rep. NBER Working Paper No. 31422, July 2023. URL: `http://www.nber.org/papers/w31422`.

[3] Infocomm Media Development Authority. *Privacy Enhancing Technologies (PET) Sandbox*. 2022. URL: `https://www.imda.gov.sg/How-We-Can-Help/Data-Innovation/Privacy-Enhancing-Technologies-Sandbox`.

[4] The White House. *US and UK launch innovation prize challenges in privacy-enhancing technologies to tackle financial crime and public health emergencies*. 2020. URL: `https://www.whitehouse.gov/ostp/news-updates/2022/07/20/u-s-and-u-k-launch-innovation-prize-challenges-in-privacy-enhancing-technologies-to-tackle-financial-crime-and-public-health-emergencies`.

[5] *UNECE Project on Input Privacy Preservation: Final Report*. Tech. rep. UNECE High-Level Group for the Modernisation of Official Statistics, 2023. URL: `https://statswiki.unece.org/x/mQCQFw`.

[6] Fabio Ricciato, Albrecht Wirthmann, and Martina Hahn. “Trusted Smart Statistics: How new data will change official statistics”. In: *Data & Policy* 2 (2020), e7. DOI: `10.1017/dap.2020.7`.