



CyberPeace
Institute

CyberPeace Institute's Approach to Responsible Use of Artificial Intelligence

SEPTEMBER 2023

CyberPeace Institute's Approach to Responsible Use of Artificial Intelligence

The mission of the CyberPeace Institute is to defend the most vulnerable in cyberspace. To do this, we provide free cybersecurity for those in need, track malicious actors, collect cybercrime evidence, and advocate for stronger regulations to protect security, dignity, and equity in cyberspace. However, when an organization is committed to such an essential mission during successive crises and constant disruptions (legal, technological), there is an inherent risk of creating a blind spot in its operations. As our work involves examining current and past attacks, investigating known vulnerabilities, and monitoring the (often poor) application of national or international law, it can make our thinking too linear: methodical, sequential, building from past experience, but missing to grasp the exponential reality of the world we live in.

To mitigate this risk, we rely on our cyberpeace disruption monitoring capability, where we monitor technological innovation, modifications in digital usage, shifts in the information landscape, and the negotiation and/or implementation of digital regulations in a volatile geopolitical context. We monitor how the convergence of these disruptions impacts the cyber threat landscape and creates potential new harm to vulnerable communities. We incorporate this knowledge into the design and delivery of all our products and services. In this context, we are closely monitoring the shifts in the technological and regulatory landscape of Artificial Intelligence (AI), particularly the recent leap of generative AI, which is converging with several legal initiatives across the globe, such as the EU AI Act, and the U.S. Blueprint for an AI Bill of Rights.



We approach this convergence from two perspectives. On one hand, we seek to understand how the shift in the AI landscape impacts our operations regarding cyberpeace, including how we provide free cybersecurity, analyse cyberattacks, and advocate for change. But more importantly, we reflect on our own usage of AI. As a data-centric organisation, where all our products and services depend on data collection, processing, analysis, and presentation, we consider it crucial to publish our internal position on the use of AI, just as we have a public position on how we collect and process data.¹ It is clear that AI can accelerate all our operations, from vulnerability scanning and cyberattack tracking to data triage, big data visualisation, and policy analysis. However, these accelerations should be understood, assessed in the context they happen, and regulated. It would be ironic for an organisation to call for regulations and accountability while considering its technology usage as a free ride.

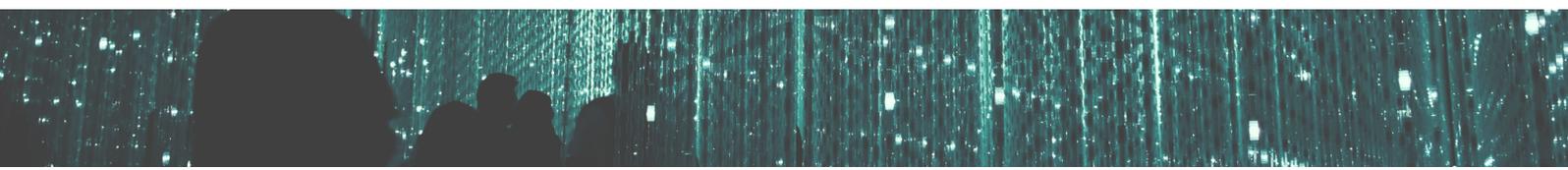
The following principles on the usage of AI guide our daily work: data collection and processing, knowledge production, talent management, stakeholder engagement, and critical assessment of partnerships. Ultimately, they guide three fundamental activities of the CyberPeace Institute: how we create and grow expertise, how we generate knowledge, and how we make decisions. These principles will also shape our public stance on AI usage, as well as AI regulations, norms, and standards.

These principles exist in a fast-moving world. They are a tool for the Institute to live up to its core values of independence, neutrality, integrity, inclusiveness, and transparency when using technology. Like everything else, they are subject to change and, over time, will be disrupted. Meanwhile, they serve as the foundation for how we envision peace and justice in cyberspace and how we commit to using technology ethically and responsibly.²

Stephane Duguin, CEO, CyberPeace Institute

¹ CyberPeace Institute. 2023. "Frequently Asked Questions: Data & Methodology | CyberPeace Institute." Cyber Attacks in Times of Conflict. 2023.
<https://cyberconflicts.cyberpeaceinstitute.org/faq/data-and-methodology>.

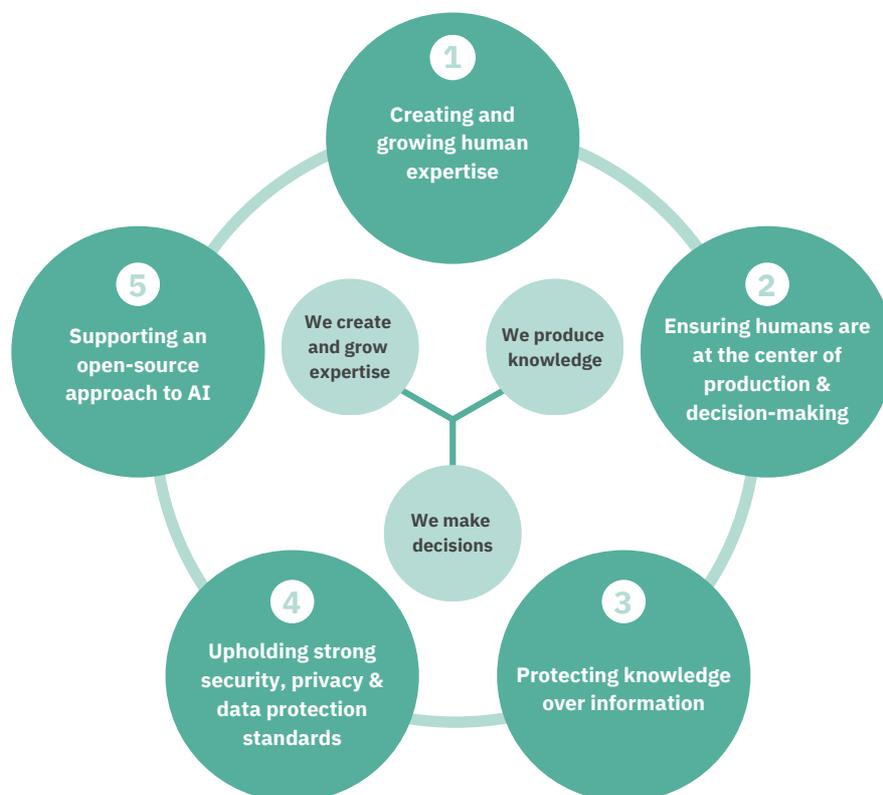
² While these principles provide an overarching framework, detailed implementation and specific actions will be further guided by internal policy guidelines of the CyberPeace Institute.



AI Usage Policy

We recognise that responsible adoption of AI requires informed, careful, strategic thought to improve the ability of our staff to carry out their human-centered work. To guarantee that we responsibly use AI, we identify five principles to guide our usage of AI at the CyberPeace Institute which we will develop below.

For the sake of this communication, we define AI systems in accordance with the OECD definition as “[...] a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.”³ This communication focuses on large language models (LLM), and machine and deep learning (ML/DL) models.



³ OECD. 2022. “Recommendation of the Council on Artificial Intelligence.” OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>.

Creating and growing human expertise

We commit to welcoming and growing the human expertise needed to achieve cyberpeace. Beyond the Institute, we commit to supporting the increase of global human expertise in protecting the most vulnerable in cyberspace. Therefore, we will never seek to blindly replace human-managed processes with AI. We will always seek to welcome and grow collaborators to make the best impact by using AI. This is especially important for entry-level positions, as we don't forget that today's experts were the ones performing somewhat repetitive or boring tasks yesterday. Safeguarding entry-level jobs is crucial for new talents to acquire knowledge and experience within the field of cybersecurity. We commit to increasing AI literacy within the CyberPeace Institute by training our colleagues on how to use AI tools responsibly and building awareness on the risks and opportunities of the AI ecosystem.

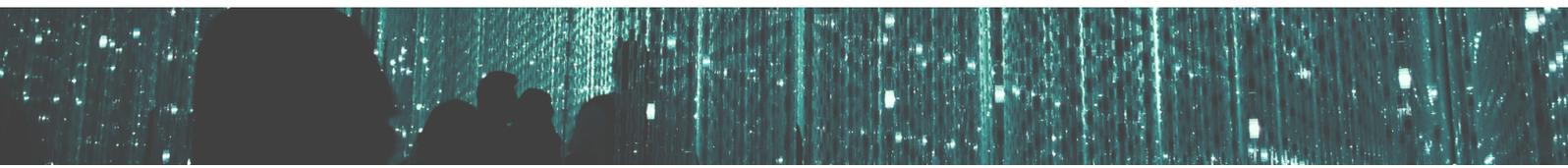
Ensuring humans are at the center of production and decision-making

We commit to never using AI to autonomously produce knowledge or make decisions. We value the human expertise we foster and invest in, which is why all our products and services result from the specialised expertise of our staff.

In all of our activities, from providing cybersecurity assistance to researching policies, we collect data, transform this data into information, and use this information to produce knowledge or make decisions. In this value chain, we acknowledge the benefit of AI to support data gathering and information synthesis. However, knowledge production and decision-making will never occur through the sole use of AI, but rather will always result from human expertise. We commit to human oversight in each step of our processes, and through this human-centric approach, we want to ensure accountability for our work and final products.

Protecting knowledge over information

We commit to the production of high-quality knowledge, based on human expertise. The editorial ability and speed at which generative AI models can produce significant amounts of content pose a great challenge to the information environment. We already live in an age of information abundance, and flooding the space with more AI-generated content will make it harder to find, consume, and process reliable, actionable, and high-quality knowledge.



Recognising this risk of generative AI diluting the information space, we refuse to participate in the information overload by ensuring that we only publish high-quality, human-created content. Our commitment to safeguarding human-produced knowledge is our support to a higher quality knowledge environment across the internet.

Upholding strong privacy and data protection standards

We believe that the existing data protection and privacy laws apply to the use of AI. This includes both external laws such as the Swiss Federal Act on Data Protection or the EU's General Data Protection Regulation as well as our internal [privacy policy](#). We commit to upholding the highest cybersecurity standards and protect our organisation's data and assets, in line with our internal cybersecurity policy.

Supporting open source approach to generative AI

We commit to supporting the open-source approach toward generative AI projects. In the same way we recognise that open source projects are fundamental for an open, interoperable, free, and secure internet, we consider the open source approach to be most apt to democratise access to efficient AI models and nurture collaborative development of AI. Open source development enables a cross-disciplinary approach to research and experimentation. Most importantly, it offers an effective way to ensure transparency and explainability in AI as it enables decentralised audits, as well as oversight and scrutiny of new AI models and the related datasets to train these models. We commit to ensuring that AI regulation will safeguard such an open-source model.

Whilst we are not solely using open-source AI technology, we do publish our open-source tools on open-source platforms such as GitHub and HuggingFace.

We commit to ensuring that these principles regarding the responsible use of AI continue to be upheld in the CyberPeace Institute. We pledge to use AI in a way that safeguards human expertise and quality knowledge, whilst guaranteeing the safety of all those connected and involved with the organisation. Should we use AI in the production of any products or services, this will always be referenced and documented.

Published on September 20th, 2023

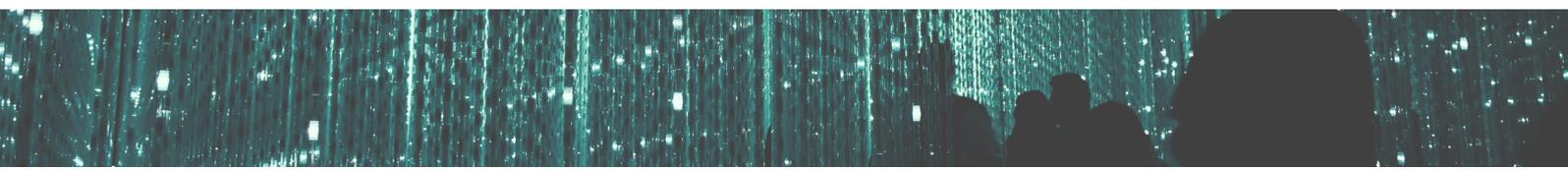




Image Source: Unsplash / Robynne Hu

© 2023 CyberPeace Institute. All rights reserved

