Other Publication

A Neutral Hub for AI Research / Ein neutraler Hub für KI-Forschung / Un hub apolitique de recherche en IA

Author(s):
Fischer, Sophie-Charlotte; Wenger, Andreas

ETH Library

CSS
ETH Zurich

A Politically Neutral Hub for Basic AI Research

In the international policy discourse, artificial intelligence (AI) is frequently discussed narrowly in terms of a new technology race between great powers. However, international science diplomacy can make an important contribution to promoting the manifold opportunities enabled by AI, while mitigating some of the associated risks.

By Sophie-Charlotte Fischer and Andreas Wenger

Artificial intelligence is driving a comprehensive transformation of the economy, society, and the state. Although the origins of AI research go back to the 1950s, AI has only become a key issue on the international political agenda in the past few years. The reasons for this are mainly technical and economic in nature. The combination of three factors – the rapid increase of computing power, the vast increase in data, and the optimization of algorithms – have facilitated a new wave of progress in AI research and development over the past 15 years. Consequently, more and more AI technologies developed in company and university laboratories have made their way into everyday practical use. In the process, AI's potential economic and social benefits are becoming apparent – but so are its considerable risks.

Large global technology companies, especially in the US and China, have made big strides in accumulating and developing knowledge, data, and applications relating to AI. Societal and state actors, however, are lagging behind. Tech companies have assumed this leading role thanks to the excellent framework conditions that they provide for the develop-

ment and commercialization of AI, including large computation power, vast datasets, and considerable financial investments in research and development (R&D). When combined with attractive salaries, flexible working conditions, and space for creative work on interesting challenges, this is the kind of environment that allows companies to attract the world's leading AI researchers. At the same

Key Points

▮ State actors increasingly view AI as a strategic resource and try to influence the innovation process and the proliferation of AI technologies.

▮ In the context of increasingly competitive political rhetoric, the risks associated with as-yet immature AI technologies are growing.

▮ What is needed is a politically neutral hub for basic AI research, committed to the responsible, inclusive, and peaceful development and use of the new technologies.

▮ Switzerland has a dynamic AI ecosystem and is also well positioned politically to serve as host state for AI governance initiatives.

time, the concentration of AI R&D in a few large corporations is increasingly undercutting the competitiveness of public research institutions and smaller companies.

Meanwhile, state actors anticipate that AI as a dual-use technology will have far-reaching effects for the global distribution of economic and military power. Increasingly, AI is regarded as a strategic resource and treated as such. More and more states are developing national AI strategies aimed at securing an advantageous position in the field. Due to the enormous economic potential of AI and its relevance for security policy, states are increasingly trying to influence the innovation process and proliferation of the new technology. They are providing extensive funding for the training of AI experts and R&D, and forming partnerships with technology companies to acquire and use AI for public purposes. At the same time, governments are trying to shield their national AI resources from unwanted external influence and to prevent a transfer of AI technologies to competitors – even though the effectiveness of such measures is in dispute. Examples include efforts to protect the national technology and industry base from some foreign direct investments, or deliberations on limiting exports of specific AI technologies to competitors.

The concentration of AI resources in private tech companies and the attendant dynamics of political competition between great powers create challenges for societies and regions that are less advanced in the AI field. Economically, they run the danger of becoming overly dependent on global technological oligopolies; politically, they might become more and more dependent on the political decisions of other actors. At the same time, fundamental technological and societal risks associated with AI may become secondary. In the context of an intensifying technology competition, the perception of an existing first-mover advantage in certain strategically important areas may trigger a "race to the bottom". If the future of this transformative technology is determined by a few big actors racing against the clock, there is a risk that the implementation of immature AI technologies will lead to accidents and will have unintended negative social impacts.

These dynamics threaten to further exacerbate the existing inequalities between societies and regions. The development of new technologies should not be shaped by a few dominant economic and political actors. To anticipate and reduce risks and make the benefits of AI accessible to all, policymakers should include voices from the economic, societal, and political spheres, and from every region of the world. They should establish effective channels for international cooperation in order to ensure that the mid- and long-term development and use of AI is transparent, inclusive, responsible and sustainable.

Science Diplomacy for international AI Cooperation

The link between science and diplomacy is a promising instrument for counteracting some of the technical and societal risks associated with AI. Science diplomacy is ideally suited for building bridges between societies and developing joint strategies for overcoming international challenges. Scientific cooperation facilitates a constructive dialog between societal, economic, and state actors because it is oriented according to the values and methodological standards of science. Researchers strive to engage in cooperation for the purpose of gaining new scientific insight, in the process supporting mutual understanding and the establishment of personal and institutional links across political and ideological boundaries.

In retrospect, we can see that science diplomacy has already made crucial contributions to fostering mutual understanding in international relations. Scientific mega-projects such as the International Space Station (ISS) or CERN (the European Organization for Nuclear Research) have proven to be effective mechanisms in facilitating international political cooperation.

Another reason science diplomacy seems promising as an approach for resolving global AI challenges is that the AI research community is very international in nature and has established an exceptionally open culture. Many AI researchers publish their results on the arXiv online platform, where they are freely accessible to other interested users for further research and development. They have harshly criticized state efforts to limit transnational research and scholarly exchange in the AI sphere.

A Neutral Hub for AI Research

We propose, as one element of a global AI governance architecture, the creation of a politically neutral, international, and interdisciplinary hub for basic AI research that is dedicated to the responsible, inclusive, and peaceful development and use of AI. The nexus between science and diplomacy can contribute to creating a sustainable mechanism for international cooperation in the field of AI, and can make the potential of AI accessible to as large a part of the world population as possible. Launching such an initiative would indicate that the development trajectory of AI is not predetermined, but can be shaped in an international framework. Moreover, the initiative would support the understanding that the development and application of AI need not be a zero-sum game. The vision of an international research platform for AI has already been suggested in a number of contexts, for instance at the AI for Good Summit or in discussions within the framework of the OECD. Our aim, in the following remarks, is to outline a potential version of such an initiative and to encourage a public debate about the opportunities and challenges associated with it, as well as its possible design. Additionally, we will discuss the special role that Switzerland could play in the implementation of such a proposal.

An international hub for AI research would have four main functions in counteracting some of the emerging areas of tension at the international level:

The first is the establishment of a globally competitive, international, integrative platform for fundamental

research in the field of AI. The hub should be an appealing laboratory that attracts the best AI talent from around the globe to facilitate basic AI research across national boundaries and to cultivate AI as a global good. The research framework of this platform would cover a broad spectrum of methodological approaches in AI. Though many current advances in AI are based on deep-learning, alternative paradigms and researchers from related disciplines such as neurosciences should be involved, as they can contribute new insights. This initiative would adhere to the highest scientific and ethical standards and offer an attractive work environment for the best AI scholars.

Secondly, the hub would serve as a platform for researching, reflecting upon, and managing the technical and societal risks associated with AI. These include, for instance, technical safety risks associated with AI and the transparency of algorithmic decision-making processes. The investigation of these risks should also include researchers from the humanities and social sciences.

Third, the AI hub should contribute to the development of norms and best practices regarding the applications of AI. The international interdisciplinary community of researchers should participate in future AI governance processes and bodies in a consultative role. This could help prevent a few powerful actors from dominating the formulation of AI norms.

Fourth, the initiative should serve as a center of learning and further education for AI researchers. Master and PhD students who complete part of their education at the AI hub would serve as multipliers as they continued their careers in other organizations. Via research and training partnerships, the AI hub would help strengthen the AI community in this area. Researchers from various disciplines and countries could also collaborate at the hub on

the development of AI applications designed to address UN Sustainable Development Goals. A decentralized approach might also be fruitful. This would include the establishment of project-based subsidiaries of the hub in other regions of the world, allowing for the development of problem-oriented AI applications with local expertise and in partnership with relevant societal groups.

Governance of the AI Hub

In principle, membership in this international AI hub should be open to all states. In all likelihood, however, this large-scale scientific project would be initiated by a "coalition of the willing". Securing the participation of a trans-regional group of likeminded states from the outset is vital. This would avoid the appearance of the initiative being the project of a select group of states seeking unilaterally to advance the competitiveness of their own region, which might prevent other countries from joining. The AI hub can fulfill its purpose if it is accepted by the international research community and constitutes an important node in a global network for safe and responsible AI.

The hub would make the results of its research available to member states as a collective good. Much as with CERN, the research groups of the AI hub should be able to cooperate with partner laboratories around the globe. Furthermore, all member states would benefit from a jointly created infrastructure for AI research functioning on a level that the individual countries are unable to provide on their own. Financial contributions from member states should be aligned with their respective economic resources and can also be made as in-kind contributions. This ought to lower the barrier to participation for less wealthy states.

There are several possibilities as to how the AI hub could be integrated with the UN. Much like the International Atomic Energy Agency (IAEA), the hub could be established as an autonomous scientific-technical organization that is linked to the UN via a cooperation agreement and that reports regularly to the UN General Assembly and the Security Council.

A Role for Switzerland

For a variety of reasons, Switzerland is particularly well positioned to take a lead role in advancing the vision of an international AI research hub and to serve as host state in its implementation. As a UN location, a non-EU member, and as one of the world's most globalized countries, Switzerland could play an important bridging role to ensure that the hub has a trans-regional orientation from the very start. With its political neutrality, its stability, its self-reliance, and its experience in helping shape multi-stakeholder pro-

Further Reading

AI Governance: A Research Agenda *Dafoe, A., Center for the Governance of Artificial Intelligence, University of Oxford, 2017.*
This report proposes a comprehensive new research agenda for AI governance.

Concrete Problems in AI Safety *Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané, 2016.*
This paper explores different research problems around ensuring that cutting-edge machine learning systems operate as intended.

The CERN Community; A Mechanism for Effective Global Collaboration? *Mark Robinson, Global Policy, Vol. 10 (1), 41−51, 2018.*
This article examines by the example of CERN why science mega-projects have been particularly effective in enabling international collaboration.

cesses, Switzerland would make a credible host for a global AI research platform. Switzerland's special historic effort to ensure that only civilian, not military research is conducted at CERN gives it further credibility in this regard.

Additionally, Switzerland offers an advantageous combination of economic and scientific conditions that are essential for the implementation of such a project. The country already has a dynamic ecosystem in the AI field and related disciplines, including excellent technical universities such as ETH Zurich and the École polytechnique fédérale de Lausanne (EPFL), a lively startup scene, and global corporate technology leaders. With the joint participation of the federal administration, the research sector, and industry in the initiative, national capabilities would be better integrated and would also have access to an even better international network, due to their critical mass. As the home of a politically neutral AI hub, Switzerland, with its high standard of living and its central location in Europe, could become an attractive alternative to locations such as Silicon Valley and other current destinations of choice for the world's top AI researchers.

At the international level, Switzerland has the opportunity to promote itself as one of the world's leading research and innovation locations, as an important facilitator of international cooperation, and as a bridge-builder at the intersection of peace support policy and foreign technology relations. Domestically, the initiative is an opportunity for Switzerland, given its strong international ties and its dependency on foreign trade, to make a credible contribution to the development of global AI norms. This contribution would reflect Switzerland's values and interests and blend the knowledge and skills of an innovative public administration, an agile industry, and world-class research community. Advocating a politically neutral, international, and interdisciplinary hub for AI would be a pioneering move that combines Switzerland's engagement on behalf of peace, which is based on the nation's own history, with the technological capabilities of a highly developed small state.

Sophie-Charlotte Fischer is a PhD candidate at the Center for Security Studies (CSS).

Andreas Wenger is professor of International and Swiss Security Policy at ETH Zurich and Director of the Center for Security Studies (CSS).