January 2018 ·
Ben Scott, Stefan Heumann and Philippe Lorenz

Artificial Intelligence and Foreign Policy

Stiftung Neue Verantwortung

Think Tank at the intersection of technology and society

Executive Summary

The plot-lines of the development of Artificial Intelligence (AI) are debated and contested. But it is safe to predict that it will become one of the central technologies of the 21st century. It is fashionable these days to speak about data as the new oil. But if we want to "refine" the vast quantities of data we are collecting today and make sense of it, we will need potent AI. The consequences of the AI revolution could not be more far reaching. Value chains will be turned upside down, labor markets will get disrupted and economic power will shift to those who control this new technology. And as AI is deeply embedded in the connectivity of the Internet, the challenge of AI is global in nature. Therefore it is striking that AI is almost absent from the foreign policy agenda.

This paper seeks to provide a foundation for planning a foreign policy strategy that responds effectively to the emerging power of AI in international affairs. The developments in AI are so dynamic and the implications so wide-ranging that ministries need to begin engaging immediately. That means starting with the assets and resources at hand while planning for more significant changes in the future. Many of the tools of traditional diplomacy can be adapted to this new field. While the existing toolkit can get us started, this pragmatic approach does not preclude thinking about more drastic changes that the technological changes might require for our foreign policy institutions and instruments.

The paper approaches this challenge, drawing on the existing foreign policy toolbox and reflecting on the past lessons of adapting this toolbox to the Internet revolution. The paper goes on to make suggestions on how the tools could be applied to the international challenges that the AI revolution will bring about. The toolbox includes policy making, public diplomacy, bilateral and multilateral engagement, actions through international and treaty organizations, convenings and partnerships, grant-making and information-gathering and analysis. The analysis of the international challenges of the AI transformation are divided into three topical areas. Each of the three sections includes concrete suggestions how instruments from the tool box could be applied to address the challenges AI will bring about in international affairs.

Economic Disruption and Opportunity

The driver of AI technology development is primarily economic. AI has the potential to reshuffle winners and losers in global markets. Without question, positioning for domestic economic interests in global AI markets as well as an AI-inspired development program will be important objectives for foreign policy leaders. However, we see the major strategic priorities for economic policy planners within foreign ministries as focused elsewhere. Because market forces are likely to move faster than policy-making, the focal points for foreign ministries are more likely to be rooted in risk management on two

major issues: 1) concentration of economic power; and 2) labor market disruption. Foreign ministries should re-tool their observation and reporting tasks to include careful monitoring of developments in AI technologies and markets. This data might be factored into risk assessments with respect to regional instability, migration, and trade. A second area of activity will be initiating international dialogue with like-minded partners to prepare the groundwork for collective action around common interests, for example on regulatory policy with respect to AI.

Security and Autonomous Weapons Systems

Among the many ways that AI might transform our societies, none have the urgency carried by the prospect of autonomous weapons. Once the stuff of science fiction, a future featuring robotic killing machines and algorithms empowered to deliver lethal force is closing fast. The top priority in this area is updating arms control and non-proliferation strategies to deal with an escalating AI arms race. In particular, this means aligning major powers around common policies (such as limitations on offensive capabilities) and working together in the common interest of guarding against these weapons falling into the hands of terrorists. This work should be accompanied by significant public diplomacy to establish moral red lines and convene influential stakeholders across sectors to contain the threat of AI weapons. In addition, there is much work to be done evaluating the potential threats of AI in hard power as well as in disinformation campaigns. There is too little understanding in our ministries about how these technologies work, which players in which markets offer weaponized AI as a product, and how we might be able to push back against them.

Democracy and Ethics

The job of foreign ministries in most liberal democracies includes two straightforward and related tasks that reflect the values of open societies. The first is to promote and strengthen democratic institutions that protect social equality and representation around the world. The second is to pursue a (human and civil) rights-based system of governance, commerce, and security in the international community. The diplomatic and development agenda surrounding the Internet has demonstrated for years the tensions between security and freedom implicit in ever more connected societies. AI will heighten this tension by supercharging surveillance and censorship capabilities. Even as these technologies enable new opportunities for free expression, civic activity, and social progress, they also raise the unwelcome possibility of deepening existing social discrimination. The challenge for foreign policy will be to promote a positive agenda in the face of these risks – leveraging grant-making, communications, and multi-lateral policy engagement to pursue rights-based goals. In their own practice, ministries that embrace data-driven AI tools for development aid projects (a likely, and

potentially fruitful, prospect for the medium term) should keep the problem of bias front of mind.

Grand theory about technology-driven change at the global level must be instrumented through institutions. And we recognize that these institutions operate under constraints – political, budgetary, bureaucratic, and human resources. Consequently, we opted to present a pragmatic proposal for the foreign policy of AI that leverages the existing tools of diplomacy while working towards more systemic adaptation in the future.

This report is drawn from an expert, multi-stakeholder workshop conducted by the Stiftung Neue Verantwortung in Berlin (September 2017) in partnership with the German Federal Foreign Ministry and the Mercator Foundation. The authors bear the responsibility for this text but wish to acknowledge with deep gratitude the contributions of all participants and the financial support of the Mercator Foundation.

Introduction

Just ten years ago, it was "cutting edge" in foreign policy circles to be focused on the role of the Internet in international affairs. The potential impact of connection technologies on foreign relations was an emerging issue for policy planners. Smart phones were new. Facebook and Twitter were interesting new companies. And the ubiquity of connectivity had not yet become a global phenomenon. When then-Secretary of State Hillary Clinton gave a major speech on Internet Freedom in January of 2010 and made the issue a priority for the State Department, it was a bold new intervention – and considered by many foreign policy experts as a distraction from more serious matters. Very few foreign ministries even staffed technology experts, much less at senior levels. Then came Wikileaks, Stuxnet, the Arab Spring, Gezi Park, and the consistent presence of digital communications as a factor in social and political movement formation. The diplomatic community hustled to catch up to events. Not only was little of this predicted by reporting officers in embassies around the world; there were very few institutional structures in our ministries to develop policy or implement programmatic work. Today, this has changed – accelerating again after the Snowden revelations. Most major capitals have cyber-units in their foreign ministries. Cyber is a hot topic in foreign policy think tanks and research institutes. And the role of the Internet in international economic, security, and social policy is recognized as important even if not fully understood.

The experience of integrating technology-focused knowledge and skills into our diplomatic practice was not simple. And for most organizations it has been unevenly implemented and has yielded mixed results. Most of the changes have been pragmatic, incremental reforms. New practices to tackle digital technologies that represent major changes in diplomatic work are few and far between – rarely matching the level of transformation in society at large. Nonetheless, the process of adaptation to technological change must become a part of standard operating procedure and begin to stretch the conventional pace of institutional reform.

The cycle of technological development is now turning again. The new, transformative, general purpose technology is Artificial Intelligence (AI).[1] AI is a term that means different things to different people. But we will use it here to mean technologies that enable machine learning, natural language processing, deduction through vast data-computational power, and ultimately, automated decision-making in robotics or software that can substitute for tasks once performed exclusively by human action and judgement. The algorithms of AI have surged up the development curve at a rate that few predicted. Witness for example the rapid advances in autonomous vehicles that many experts considered impossible only a few years ago. And while not everyone fears the imminent arrival of the Singularity – the idea that non-biological intelligence will one day surpass human ability and transform civilization – the near-term capabilities of AI are jaw-dropping.

Advances in AI-powered drone technology will soon put low-cost, precision weapons in the field to conduct armed conflict without human risk to the attacking force. Early forms of these weapons are already in hands of non-state actors. ISIS is already reportedly using modified commercial drones to attack Iraqi tanks.[2]

Companies are planning to upgrade to the so-called "lights out" factory – where robots work 24/7 to manufacture, package and ship products without human supervision. Amazon has reduced its "click to ship" time from 60-75 minutes to 15 minutes with robot labor."[3]

The precision of AI-driven facial recognition software has advanced dramatically, permitting security agencies extraordinary new powers of surveillance. To demonstrate the foreboding potential with the banal, Chinese police have begun to display the names of jaywalkers on huge roadside billboards.[4]

The plot-lines of AI development remain far from clear at this point. But it is safe to predict that it will become one of the central technologies of the

1 Brynjolfsson, E. Rock, D., Syverson, C., 2017. *Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*. National Bureau of Economic Research. Available Here.

2 Warrick, J., 2017. *Use of weaponized drones by ISIS spurs terrorism fears*. The Washington Post. Available here.

3 McKinsey Global Institute, 2017. *Artificial Intelligence. The next Digital Frontier?* McKinsey Global Institute. Available here.

4 Chin, J. & Lin, L., 2017. *China's All-Seeing Surveillance State Is Reading Its Citizens' Faces*. The Wall Street Journal. Available here.

21st century. It is fashionable these days to speak about data as the new oil.[5] But if we want to "refine" the vast quantities of data we are collecting today and make sense of it, we will need potent AI. This is true for sought-after technical breakthroughs in medical diagnostics, sensors in the industrial internet, and road-scanning by autonomous vehicles. Incredible opportunities for human progress may come with AI. But AI will also challenge fundamental ethical, economic, and security institutions of our time. We might harness the power of these tools to advance human progress a giant step forward. But we may also suffer grave calamities. How government manages and responds to these technologies will play a major role in charting the course.

The imminent transformations of AI intersect with conventional foreign policy issues in fundamental ways. At the highest level, it is the impact on the balance of global power. The potential that AI brings to advance national economic and security interests has triggered a heated competition among governments to gain a strategic advantage. China's national AI strategy shows how seriously governments take this technology, placing major bets on the the future of this industry. In a recent speech, Russian President Vladimir Putin stated bluntly that the country that gains an edge in AI "will be the ruler of the world."[6]

This paper seeks to provide a foundation for planning a foreign policy strategy that responds effectively to the emerging power of AI in international affairs. There is a spectrum of possible reform options, ranging from aggressive to pragmatic. An aggressive approach would require a far more decisive move towards reshaping our foreign policy institutions than we have seen in the cyber-strategies in response to the Internet as a global change vector. It would mean a major shift in human resource acquisition, training, pathways of promotion, and the very definition of what it means to work as a diplomat in the information age. We believe this is ultimately necessary to meet the needs of the decades ahead, and this paper points out some elements of this forward leaning strategy.

However, we have chosen primarily to sketch a pragmatic approach to the foreign policy of AI. The developments in AI are so dynamic and the implications so wide-ranging that ministries need to begin engaging immediately. That means starting with the assets and resources at hand

5 The Economist, 2017. *The world's most valuable resource is no longer oil, but data.* The Economist. Available here.

6 AP News, 2017. *Putin: Leader in artificial intelligence will rule world.* Associated Press. Available here.

while planning for more significant changes in the future. Many of the tools of traditional diplomacy can be adapted to this new field. While the existing toolkit can get us started, this pragmatic approach does not preclude thinking about more drastic changes that the technological changes might require for our foreign policy institutions and instruments.

We begin with a review of lessons learned within our foreign policy institutions from the application of statecraft adapted to the disruptive power of the Internet. A few broad observations generic to any technology change offer important guidance about optimizing the strategy for tackling AI. From this basis, we tick through a short list of the foreign policy "toolbox" and suggest the outlines of how adaptation begins with what we know best. We do not mean to suggest that there should be no effort to reach for new ideas and even new concepts of 21st century diplomacy – and we encourage that as well. But for the near term, large diplomatic institutions have a delimited political remit and a relatively fixed toolkit of practices that may be applied to any given problem. We do not propose a radical transformation. On the contrary, we indicate how each of these conventional methods may fit with emerging issues of AI. By setting an institutional framework of how foreign policy practice may change, we offer the realistic context for policy planning that will be shaped by the issue analysis that follows.

The second half of the paper is a specific review of three major areas of intersection between AI and foreign policy: 1) Economic Disruption; 2) Security and Autonomous Weapons; and 3) Democracy and Ethics. It is not intended to be a comprehensive survey of all the issues that AI technologies may raise for diplomatic and development practice. It is rather an intentional prioritization of these three categories with a preliminary review of the central issues in each. For each area, we offer a definition of the foreign policy problem(s), a strategic objective for diplomatic practice, and a set of initial policy proposals that may guide planning. The fundamental purpose of this paper is twofold – to offer an overview of an under-attended, emerging problem in foreign policy (AI technologies) and to provide a jumping off point for policy planning in diplomatic institutions.

Laying the Foundation for an AI Foreign Policy Agenda

How should foreign ministries respond to the wave of changes that are coming? Of course, we do not begin from zero with this exercise. The starting points are the lessons learned from how the foreign policy community responded to the complex influence of the Internet on international relations and foreign affairs over the last decade. There are four broad structural conclusions we can draw from this assessment that are worth highlighting here.

The first and simplest lesson is about the internal organization of our diplomatic institutions. We will have to be faster, more experimental and risk-tolerant in our methods to test different approaches to problem solving. And we must be more ambitious in our efforts to integrate technology knowledge into the conventional organizational units of our institutions. Following this assessment to its logical conclusion would require very substantial institutional reorganization that few ministries are prepared to undertake. But even a pragmatic strategy, to be effective, will have to stretch the boundaries of what we have done so far on Internet-era statecraft. For example, it was not enough to have a small office with a few people handling all things "cyber." It will not be enough to create a special office for AI. The changes profiled in this paper speak to a systemic implications that will alter many different areas of foreign policy work – from economics to security to democracy promotion. The knowledge, skills, and process re-engineering needed to respond effectively will need to be distributed across our institutions. These operations may still be coordinated by a centralized team of experts, but the need cannot be met simply with a "special envoy". Of course, this systemic approach also means that if we prioritize AI and develop new competencies accordingly, we will have to de-prioritize other issues with decreasing relevance. Those are often the more difficult conversations but they are essential for effective and successful adaption.

The second lesson is that an effective response will be a multi-stakeholder affair with the ministry as an important hub in a network of actors that includes private companies, research institutions, civil society organizations, the media, and of course, other government agencies with adjacent remits. An effective collaboration is essential to acquire knowledge quickly, to identify the most useful interventions, and to avoid duplicating effort, working at cross-purposes or simply repeating the mistakes of others. By establishing a broad base of collaborators, we can best draw out existing

competencies. Many of these strategies may already be present in the ministries. For example, many of the AI policy issues related to security have parallels in the arms control efforts of the Cold War and post-Cold War periods.[7] The partnerships with the technology industry and human rights organization in promoting Internet freedom is another useful example.

The third lesson that we can take from our experience handling the foreign policy response to the Internet is to build adaptation into our method of problem solving. This does not mean inventing new tools of diplomacy – that toolbox (described below) is relatively fixed. But it does mean avoiding the tendency of all bureaucracies to gravitate towards concretizing general rules and procedures to handle the specific problems that arise from a topic area. The pace of change for AI is simply too fast and requires conscious effort at adopting work structures that include persistent review and revision. We propose here a simple, cyclical rubric of planning and implementation with these steps: 1) knowledge acquisition; 2) problem definition and prioritization; 3) developing and testing competing proposals for new policy or programs based on core tools; 4) pilot implementation projects; 5) lessons learned that inform a new round of knowledge acquisition as the technology and its global impact evolves. This iterative problem solving process is axiomatic in the software industry that drives the AI market. And many of its features can be usefully applied in policy development as well.

The fourth lesson is that we should expect a persistent challenge with respect to human resources – attracting, hiring, training/retaining, and promoting staff with the requisite skills. Without competency in the language of AI research and the technical advances in the commercial marketplace, it is unlikely that our diplomatic practice will suffice to meet the need. Finding and cultivating insightful analysts is normative in ministry work, where expertise in conventional foreign policy arenas is expected. But the challenge of finding and developing the needed AI experts and then integrating them into the relevant offices, divisions and embassies cannot be underestimated. The best candidates for these roles may not come through the conventional pipeline of foreign and civil service officers.

7 For examples, see the appendix in Allen, G. & Chan, T., 2017. *Artificial Intelligence and National Security*. Belfer Center for Science and International Affairs Harvard Kennedy School. Cambridge. Available here.

Toolbox for an AI Foreign Policy

Over the last decade, most foreign policy institutions have begun to engage the Internet as a disruptive force in international relations. This has been done in a variety of ways from policy planning to public diplomacy to embassy reporting and programmatic implementation projects. In most cases, the things that worked best were modifications of existing diplomatic practice. Some proved novel and innovative, such as public diplomacy on social media, but few represented a radical break from the modes of work for which large foreign policy institutions are built and funded to conduct. We believe this will hold true for addressing the change wrought by AI. While we are convinced that the most successful ministries will innovate and break convention, it is equally true that most of the work will still happen through normal modes. To that end, we offer here a sketch of the foreign policy toolbox with brief comments indicating how each of them may be instrumented to work on AI-related issues. This summary is intended as an invitation to further planning work to build upon this skeleton of a new AI foreign policy practice.

- **Policy Making** – As a first order of business, foreign ministries have an obligation to evaluate the major issues at the intersection of AI and international relations and guide the development of governmental policy positions. Particularly in the security and ethical dimensions, the trajectory of AI technologies portends the need for red lines that must be defined and articulated. We cannot shape the future of AI without first choosing objectives and paths toward them. Threat and opportunity levels must be ascertained across a range of issues, resources and staffing allocated, and institutional change set in motion.
- **Public diplomacy** – The central business of diplomacy is communication. We talk to the governments, publics, civil society and media of other nations in order to state views, refute others, and broadly advance our values and interests. Raising awareness about the implications of AI for international relations (and policy choices consistent with democratic norms) should be a new part of communications work. This should focus on priority issues and countries where change is judged likely to be near-term and high impact.
- **Bilateral and multilateral engagement** – We must start dialogue, especially with like-minded allies, to exchange views on these issues, to hear new perspectives, to examine case studies of AI in action and to test the mettle of our foreign policy responses. Ultimately, these exchanges

should be steered toward alignment around public policy choices – similar to what the international community has done with cyber-security.

- **Actions through international and treaty organizations** – The long road of international coordination and confidence building begins for AI policy in formal and informal groups of multilateral experts[8] and within the lower levels of treaty organizations that may one day contribute to norm setting or even binding international law. Short term agreements are unlikely, but the nature of some of the AI-related threats are so severe that they will likely require coordinated international action and consequences for violating the norms of the international community.
- **Convening and partnerships** – Foreign Ministries can show leadership by bringing together stakeholders from different regions and sectors to confront the challenges and opportunities of AI. Just as the early conferences on Internet Freedom resulted in strengthened alliances around a rights-based policy agenda, similar activities in AI should be initiated. This work doubles as public diplomacy, raising the reputation of the leaders of the convenings.
- **Grant-making** – There is an urgent need to build a foundation of capacity and competence in global civil society to engage the implications of an AI future. These institutions are often the recipients of international grant-making, and they can be steered towards AI if they are not already aware of the problems.
- **Information gathering and analysis** – The embassy system is designed as a distributed network of information gathering, relationship building and on-the-ground analysis with the goal of informing better policy decisions in capital cities. If we are to make AI a priority, it must become a serious part of this reporting system. AI markets and government programs must be monitored. Key leaders should be engaged by delegations. And implications for national interests – whether opportunities or threats – must be frequently flagged for intervention. In addition to monitoring the development of AI, ministries should explore using machine learning algorithms within their own systems to sort, prioritize, and find patterns within the global reporting structure of our distributed embassy system.

This blueprint for an AI foreign policy strategy is modest in scope by intent. Using this pragmatic approach, there is no revolution here in how we work – only in the topic we're working on and its demands for pace and creativity. This is applying the tradecraft of diplomacy to a new set of technological developments, but the core tools are the same. The effective adaptation of

8 For example, a group of experts has been meeting via the UN on the topic of AI weapons for a few years See, e.g. here.

general purpose foreign policy tools – if well executed – is sufficient to make progress and probably even to achieve a leadership role in the international community. To ask more of most ministries would likely be futile in the short term. However, it will ultimately be necessary if AI achieves its potential.

The key to developing successful foreign policy for AI is "effective adaption." The example of how we have dealt with the Internet in foreign policy circles is not especially inspiring. Absent an impending crisis, there is a tendency to default to conventional topics of international relations in staffing, policy, communications and programs. And we ignore the catalyst for disruption that comes from elsewhere. With the rapid emergence of AI as a change agent in economics, security and democracy, we ignore it at our peril. Now is the moment to move quickly to adapt our institutions – particularly with respect to economics, security and democracy. Pragmatic, methodical progress is the way to get the engine of change moving, but planners should have more structural reform in mind for the medium term.

Topic #1 – Economic Disruption and Opportunity

Strategic Priorities

The driver of AI technology development is primarily economic. AI has the potential to reshuffle winners and losers in global markets. The global R&D race in AI points to the importance of early market power and the probability that AI will mirror some of the winner-take-all market dynamics of the platform economy. In addition to this national competition for AI dominance, there is likely to be tension between old and new industrial development as AI makers threaten to capture the value of the traditional products into which the technology is integrated. For these reasons, a high priority in foreign policy will be advancing the interests of domestic AI business, opening markets, shaping partnerships, and guarding interests attached to intellectual property.

However, it does not have to be a Hobbesian zero-sum game. There is a strong case to predict that AI will generate substantial economic growth and prosperity for a broad set of nations and actors. Similar to what we have seen with mobile technologies, AI applications could permit some countries to achieve economic leapfrogging: skipping entire stages of development. In

areas of public service that affect economic growth – such as the quality of healthcare or education and job skill training – this will represent a welcome acceleration of progress. Low capital costs in developing AI-powered tools could enable countries to gain a significant comparative advantage in global economics and expand access to life-enhancing technologies down the socioeconomic ladder. There is an important development agenda implicit in AI market growth that places a moral responsibility on technologically advanced nations to share access to knowledge and tools that advance human prosperity.

Without question, positioning for domestic economic interests in global AI markets as well as an AI-inspired development program will be important objectives for foreign policy leaders. However, we see the major strategic priorities for economic policy planners within foreign ministries as focused elsewhere. Because market forces are likely to move faster than policy-making, the focal points for foreign ministries are more likely to be rooted in risk management on two major issues: 1) concentration of economic power; and 2) labor market disruption. Each of these pose significant threats to international economic stability as well as to national interests that cut across a variety of issues that must be addressed in foreign policy.

Concentration of Economic Power

Intensive national investment in AI research and development is designed to achieve an asymmetric advantage in new technologies that could shift the balance of global leadership. A small group of nations are currently on course to achieve dominance in critical AI technologies. The United States and China have a considerable head start. This could lead to even greater concentration of power and wealth at the top of global markets, intensifying the status quo. An AI breakthrough by a company with current market power will pose the prospect of locking in global technology monopolies. This is true not just for existing market segments determined by network effects. AI could also prove decisive in the emerging markets for the Internet of Things, autonomous vehicles, financial services, as well as weapons systems. Many nations will perceive this eventuality as a threat to sovereignty. For states that do not have corporate players in the tech oligopoly, national interests (economic, security, and social) may nonetheless be dependent on new AI. Yet, the state may have limited functional ability to control its social and economic outcomes. This tension between foreign technology interests and

national regulators is already a dynamic in global economic policy, and AI could accelerate circumstances into more frequent conflict.

The stakes of the game are well understood by global leaders. Consider current developments in the global race for AI leadership. Through its recently published *New Generation of Artificial Intelligence Development Plan*, China expects to generate 400 billion yuan ($59 billion) in AI-based economic activity by 2030.[9] The South Korean government announced an annual investment of $863 million in artificial-intelligence (AI) research over the next five years.[10] In Canada, the Trudeau government recently announced a $100 million investment in the Vector Institute at the University of Toronto – seeking to mint more AI-trained graduates than any other nation.[11] Venture investment in Canadian tech companies – powered by AI – will likely top $2 billion in 2017.[12] Meanwhile, German government research funding and contracts have created a network of AI projects intended to feed valuable innovation into the private sector.[13]

These government initiatives are all dwarfed by the private sector activities of the leading tech multi-nationals, such as the big 5 US tech firms (Facebook, Alphabet, Apple, Microsoft, Amazon) and their Chinese counterparts (Alibaba, Baidu and Tencent). McKinsey estimates that in 2016, these tech

9 Webster, G., Creemers, R., Triolo, P., Kania, E., 2017. *China's Plan to 'Lead' in AI: Purpose, Prospects, and Problems.* New America. Available here. Especially through a supercharged Chinese manufacturing sector. See China's "China Manufacturing 2025" plan: European Union Chamber of Commerce in China, 2017. *China Manufacturing 2025*. Beijing. Available here.

10 Zastrow, M., 2016. *South Korea trumpets $860-million AI fund after AlphaGo "shock".* Nature. Available here.

11 Khosravi, B., 2017. *There's An AI Revolution Underway And It's Happening In Canada*. Forbes. Available here. See also, http://vectorinstitute.ai/.

12 Financial Post, 2017. *Canadian tech venture capital funding hits eight-quarter high thanks to AI*. Financial Post. Available here.

13 Comprised of the German Research Center for Artificial Intelligence (DFKI), the relevant Fraunhofer Institutes, the Helmholtz Association of German Research Centres, the Max Planck Society, the Leibniz Association and a number of pre-eminent university departments. See Fachforum Autonome Systeme im Hightech-Forum, 2017. *Fachforum für Autonome Systeme - Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft*. Berlin. Available here.

giants alone spent between $20 billion and $30 billion on AI.[14] In the startup space, venture investment in 2016 topped $5 billion globally, with more than 60% of the money going to US-based companies.[15]

The last two years have seen an acceleration of advances towards new kinds of products – in healthcare, security, finance, and energy, to name a few.[16] These trends will likely increase pressure in regulatory debates that pit national interests versus global technology players – including data protection, cyber-security, law enforcement, and taxation. In addition, there is a race among AI's leading firms to acquire new talent – either by hiring them at high compensation or buying their companies.[17] The "brain drain" from home countries to multi-nationals (in the US and China in particular) could heighten concerns over foreign investment and acquisition of domestic technology companies and talent.

In European foreign affairs in particular, these trends will likely play out in the escalating clash with foreign technology firms. The challenge for Europe is both a question of national sovereignty as well as economic competitiveness – the former becoming a political tool to create space for the latter. The reality is that the most advanced AI-powered hardware and software solutions needed to run the industries of the future are not currently made in Europe.[18] This is not a new problem – the same companies that lead on AI also provide essential cloud infrastructure for data processing and hardware for enterprise class networks. But AI will underscore for Europe that this new

14 90 percent of this was spent on R&D and deployment, and 10 percent on AI acquisitions. In 2016 between $4 and $5 billion resulted from venture capital. Private equity firms are reported to have invested between $1 billion and $3 billion. $1 billion of additional investment was generated from grants and seed investments. See p. 6 McKinsey Global Institute, 2017. *Artificial Intelligence. The next Digital Frontier?* McKinsey Global Institute. Available here.

15 CB Insights, 2017. *The 2016 AI Recap: Startups See Record High In Deals And Funding*. CB Insights. Available here.

16 Robertson, S. K., 2017. *How Google Brain is making major advancements in machine learning*. The Globe and Mail. Available here.

17 Metz, C., 2017. *Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent*. The New York Times. Available here.

18 Forrester Research, 2016. *The Forrester Wave™: IoT Software Platforms, Q4 2016. The 11 Providers That Matter Most And How They Stack Up*. Forrester Research. Available here.

round of technology revolution may be the last chance to join the top ranks of global players.

Given the importance and scale of global developments in AI research, any national strategies on AI need to be informed by international analysis. This provides Foreign Ministries with opportunities to identify the strategic implications of AI market formation and its influence on the global balance of power. Policy planners across the globe will be grappling in an apparent zero-sum game to shape emerging AI markets to achieve three goals: 1) accelerate the growth of a top domestic AI industry; 2) secure partnerships between old and new industries that do not cede the primary value capture of core domestic industries to foreign tech giants; and 3) monitor/manage the acquisition of domestic technology companies, talent, and patents by foreign investors. These vectors will be shaped by each nation's desire to optimize its position vis-a-vis the new power structure of AI markets.

Labor Market Disruption

It is in the labor market that we may see the most disruptive consequences of AI, as automation displaces large segments of the low and semi-skilled workforce with robots and software.[19] Foreign policy makers must evaluate the rise of technological unemployment and job market polarization in nations and regions. We must track these phenomenon and plan for significant changes in global capital flow, labor dislocation and migration, and regional shifts in the balance of economic power.

These economic changes could shape both domestic and foreign policy agendas. Labor markets around the globe will be affected by intelligent machines substituting for manual and cognitive labor in manufacturing, transportation, and data processing.[20] In developed countries, the use of AI in software and robotics will lead to large productivity gains that will flow predominantly to capital holders.[21] In contrast, both human capital and (manual) labor – except for highly skilled AI developers and specialists –

19 McKinsey Global Institute, 2017. *A Future that Works: Automation, Employment, and Productivity.* McKinsey Global Institute. Available here.

20 Frey, C. B. & Osborne, M. A., 2017. The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, Vol. 114, pp. 254–280.

21 Avent, R. (2016). The Wealth of Humans: the Future of Work in the Twenty-first Century. St Martins Pr.

would decline rapidly in value.[22] This will put social welfare systems that are built on the taxation of labor under pressure. Increased automation and joblessness will test the endurance of social safety nets and the credibility of governments promising economic mobility to the middle class. Rising inequality will further increase populist and nationalist movements demanding justice for a displaced working class. In developed countries, this is likely to translate into further opposition to foreign migration even as the same phenomenon in the Global South triggers more population movements in search of a better life.[23]

In developing countries, automation-driven unemployment will not only increase poverty but contribute to political unrest as trends of expanding economic mobility stagnate or even reverse.[24] We have no clear answer to what will happen when large swaths of the labor force that have traditionally chosen industrial manufacturing as a way out of poverty are replaced by machines. The corporate search for cheap labor that characterized globalization in the 1990s reallocated wealth and economic opportunity with merciless efficiency – creating a new set of winners and losers around the world. Imagine a similar wave that displaces the low-wage labor success stories in Asia and Latin America with robots and automated production. Over time, the impact of technological unemployment could dwarf that produced by off-shoring manufacturing. Millions of laborers may spill out into job markets and glut the low-wage service sector with supply.[25]

There may be an uptick in high skill employment to build and operate the new technology economy. And in some cases, economic growth resulting from major productivity gains may be sufficient to cushion the transition. But given the speed of the technological change, it is unclear how fast education systems will be adapted to address emerging skills gaps in the economy. Moreover, the number of jobs created in administering AI systems is not likely to be sufficient to replace all of those displaced, even if re-skilling programs were perfect. Circumstances and outcomes will vary widely among nations, and it is clear foreign policy makers must plan for contingency.

22 Watch Sachs, J., 2017. *Robotics, AI, and the Macro-Economy.* Available here.

23 Piketty, T., & Goldhammer, A. (2014). *Capital in the twenty-first century*. Brilliance Corp.

24 Bryant, C. & He, E., 2017. *The Robot Rampage*. BloombergGadfly. Available here.

25 Avent, R. (2016). The Wealth of Humans: the Future of Work in the Twenty-first Century. St Martins Pr.

Preliminary Foreign Policy Agenda

The starting point for building a foreign policy agenda on the global economics of AI should be data gathering. Foreign ministries should re-tool their observation and reporting tasks to include careful monitoring of developments in AI technologies and markets. This data might be factored into risk assessments with respect to regional instability, migration, and trade. A second area of activity will be initiating international dialogue with like-minded partners to prepare the groundwork for collective action around common interests, for example on regulatory policy with respect to AI. And finally, foreign ministries will provide invaluable international inputs into inter-agency processes to conceive national AI strategies designed to foster domestic industry and increase competitiveness.

1. Labor Market Assessment – National studies of the potential impact of automation on labor markets have generated alarming results. But we have no statistical or categorical standards to organize and measure the phenomenon at the international level. Multilateral standard setting as well as pulling national level reporting into aggregated conclusions should be a first stage task. This work might include targeted grants into research to support development of standards and frameworks for this assessment.
2. Economic Instability Risk Assessment – Some nations and regions appear more vulnerable than others to the threat of technology-driven unemployment. Foreign policy makers should seek to create global risk profiles for economic instability to judge the probability of economic turbulence, migration flows, or political instability that might follow from rapid emergence of AI technologies in local industry.
3. Foreign Acquisition of Domestic AI Technology – An early policy priority must be an evaluation of domestic AI technology development for the purpose of establishing national interests and the policy criteria for restricting or conditioning the foreign acquisition of firms and intellectual property. Foreign policy planners bring to this interagency problem unique insight about international AI investment, R&D, and linkages between states and industry.
4. Global Data Policy Framework – Beneath the AI technology explosion is the data revolution. It is a global data economy that supplies the raw materials for machine learning and AI, from online behavior tracking to industrial sensors to the laser scanners in autonomous vehicles. One avenue of managing concentration of power in AI markets is to establish and enhance international standards of data governance that regulate

the storage and exchange of information. By creating open standards for data, such a policy could cut against the accumulation of market power.

Topic #2 – Security and Autonomous Weapons Systems

Strategic Priorities

Among the many ways that AI might transform our societies, none have the urgency carried by the prospect of autonomous weapons. Once the stuff of science fiction, a future featuring robotic killing machines and algorithms empowered to deliver lethal force is closing fast. The people in the best position to judge how near we are to this future are among those most alarmed. On July 28, 2015, an open letter was presented at the International Joint Conference on Artificial Intelligence in Buenos Aires, Argentina, calling for a ban on offensive autonomous weapons.[26] To date, this letter has been signed by over three thousand leading AI researchers, including the most renowned scientists from the leading universities in the West.[27] But the drawing of this moral red line is not a universal phenomenon. Notably, there is no equivalent approach to the ethical questions related to autonomous weapons in the Chinese discourse.[28] Russian arms manufacturers have announced plans to develop AI-powered missiles and small arms.[29] The diplomatic work to align nuclear states around a common framework of arms control does not yet extend to AI. Yet a new consensus with China, Russia and other rising AI powers on specific norms will be crucial to controlling an impending AI-

26 Various, 2015. *Open Letter on Autonomous Weapons*. International Joint Conference on Artificial Intelligence (IJCAI) 2015. Available here.

27 Such as the universities of Cambridge, Oxford, Harvard, Stanford and MIT. See Griffin, A., 2015. *Stephen Hawking, Elon Musk and others call for research to avoid dangers of artificial intelligence*. The Independent. Available here.

28 The Economist, 2017. *Code red. Why China's AI push is worrying.* The Economist. Available here.

29 Greene, T., 2017. *Russia is developing AI missiles to dominate the new arms race*. The Next Web. Available here.

weapons arms race. And of course, the potential threat from AI is not limited to nation states.

Put simply, autonomous weapons are rapidly developing into a grave national security problem.[30] Taking the right decisions, and taking them as fast as possible, is essential to winning any military conflict. Further, possessing lethal and destructive weapons that pose little risk to the lives of the operators removes a potent deterrent for armed conflict. For these reasons (among others), many believe AI technologies will revolutionize warfare. Automated killing machines cross clear ethical red lines. But just as with chemical, biological and nuclear weapons before them, that doesn't mean they won't be built and fielded. Combine this with the tactical advantages for military commands in possession of AI data processing for identification of targets, managing logistics[31], conducting surveillance, and honing training and simulation. A new arms race appears inevitable alongside a new set of dangers from terrorism.

In this context, we have identified three areas of priority interest for foreign policy planners: 1) AI weaponry and the changing balance of power; 2) Non-state terrorist activity using low-cost AI weaponry; 3) New forms of conflict focused on information and data manipulation. Of course, the responsibility to develop and apply these policies will fall across multiple government agencies in the national security system. We focus here primarily on the diplomatic components of this work.

Autonomous Weapons[32] Systems and the Global Balance of Power

Breakthroughs in AI weapons systems may create lasting, asymmetrical advantages for the world's top militaries. However, there is a strong chance

30 For anyone who seeks a deep analysis on this problem we highly recommend the recent paper by Greg Allen and Taniel Chan on Artificial Intelligence and National Security. Allen, G. & Chan, T., 2017. *Artificial Intelligence and National Security*. Belfer Center for Science and International Affairs Harvard Kennedy School. Cambridge. Available here.

31 Lofgren, J. B., Zielinski, P., 2007. *Operation Iraqi Freedom and Logistics Transformation*. USAWC Strategy Research Project. The U.S. Army War College. Available here.

32 In the US, an autonomous weapon system is defined as: "*A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation*". Department of Defense United States of America, 2012. *DoD Directive 3000.09, November 21, 2012*. Available here.

that these advantages will be fleeting and these gaps could close quickly. The cost and difficulty to produce AI-based weapons are relatively manageable compared to rocketry and nuclear technologies or fighter planes and aircraft carriers. For example, much of the AI technology needed to weaponize a drone (aerial or terrestrial) will be present in civilian-use products that may be modified for military purposes. Of course, the ease of proliferation does not make these weapons less deadly. If this is our technological future, the task of arms control will become immensely more difficult. The most dangerous AI technologies may not have a clear dual-use profile but rather appear as digital code written for entirely legitimate civilian purposes.

The consequences that access to autonomous weapons may bring to the global balance of hard power could be severe. For starters, AI weapons could serve authoritarian states as a new, relatively inexpensive option for attaining strong deterrence capabilities. We may see challenges to the balance of regional power as states move to leverage AI technology to reverse historic disadvantages vis-a-vis neighbors. The new advances may lie in the technology itself, but it may also be the perceived willingness of a country to cede lethal decisions to machines. Even if countries decide to make sure humans remain arbiters of life/death decisions (such is the law in the US[33]), such liberal states may be forced to reckon with the fact that they are putting themselves at a strategic disadvantage. Due to information processing constraints – both in terms of quantity and speed – human analysts will hardly be able to compete with AI powered decision-making. Further, there are great dangers that AI powered military systems and military decision-making will undermine existing approaches for conflict containment and de-escalation.

The implications for foreign policy leaders are grave. The institutions and treaty instruments designed for 20th century arms control and nonproliferation are not made for a world order in the midst of an AI arms race. We must prepare for more frequent and more disruptive outbreaks of violence in conflict zones as the human and financial cost of making war declines – triggering migration, economic instability, poverty, health crises and famine. Governments must reassess risk management and particularly the alarming areas of catastrophic risk management that have previously been reserved exclusively for nuclear, chemical and biological weapons.

33 Department of Defense United States of America, 2012. *DoD Directive 3000.09, November 21, 2012*. Available here. Galdorisi, G., 2015. *Keeping Humans in the Loop*. Proceedings Magazine, Vol. 141(2), p.1,344. Available here. Beard, J. M., 2014. *Autonomous Weapons and Human Responsibilities*. Georgetown Journal of International Law, Vol. 45, pp. 617-681. Available here.

If conflict becomes more frequent, there will be new challenges in post-conflict stabilization efforts, humanitarian crises, and refugees flows.

AI-Enabled Terrorism from Non-State Actors

Perhaps the greatest threat from AI weapons comes not from state actors in possession of new power, but from non-state terrorist organizations. Unlike previous military breakthroughs[34] the cost of AI weapon deployment will be low enough to fall within the scope of even unsophisticated terrorists (e.g. consider an AI guided drone carrying a chemical payload). That means the AI-based arms race will include not only national militaries but also non-state actors and asymmetric military strategies.[35] Diplomatic programs engaged in countering violent extremism and counter-terrorism will have to take these new variables into account. These kinds of weapons also represent a significant new threat for diplomatic security, protecting embassies, diplomatic personnel and citizens travelling abroad.

Adversarial examples

Adversarial examples can be used to trick a machine learning system into misclassifying an object with high levels of confidence: for example, the machine learning system interprets a stop sign as a yield sign.[36] This can be reached through "physical-world attacks"[37] for instance through manipulating the stop sign itself, and through attacking reinforcement

34 Fission / fusion bomb, intercontinental ballistic missiles (ICBM), multiple independently targetable reentry vehicles (MIRV). For details see p. 99: Bostrom, N. 2013. *Superintelligence: Paths, Dangers, Strategies*. Oxford, Oxford University Press.

35 Allen, G. & Chan, T., 2017. *Artificial Intelligence and National Security*. Belfer Center for Science and International Affairs Harvard Kennedy School. Cambridge. Available here. De Spiegeleire, S., Maas, M., Sweijs, T., 2017. *Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers*. The Hague Centre for Strategic Studies (HCSS). Available here.

36 Goodfellow, I., Papernot, N., Huang, S., Duan, Y., Abbeel, P., Clark, J., 2017. *Attacking Machine Learning with Adversarial Examples*. OpenAI Blog. Available here. Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., Swami, A., 2016. *Practical Black-Box Attacks against Machine Learning*. Computing Research Repository. Available here.

37 Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A.,Rahmati, A., Song, D., 2017. *Robust Physical-World Attacks on Deep Learning Models*. Computing Research Repository. Available here.

learning agents' algorithms[38] by providing them with malign inputs (for instance with manipulated training data).[39] Imagine a manipulated classifier causing a self-driving car to ignore a stop sign. This is dangerous. And potential attack vectors are as diverse as are use cases for machine learning algorithms. The problems connected to adversarial examples are very hard to resolve. Up to date there is no effective defense algorithm and hopes are low that there ever will be.[40] And similar to cyber attacks, attackers may have a strategic advantage over defense architectures. This is why notable researchers and institutions advocate for more research on adversarial examples.[41]

Data Warfare

It is unlikely that there will ever be another conventional military conflict that does not have components of information and cyber warfare. AI will play a central role in how these new forms of weaponry are deployed. This is of course about hacking and data exfiltration, as well as cyber-attacks aimed at causing loss of life and property. The further development of AI technologies will enhance tools of network penetration and exploitation. It may well be that AI cyber operations are simply left to engage in a constant state of attack – seeking to penetrate as many networks as possible and then lie in wait for strategic moments of exploitation. It could lead to the

38 Papernot, N. & Goodfellow, I., 2016. *Breaking things is easy*. Cleverhans-Blog. Available here. Huang, S., Papernot, N., Goodfellow, I., Duan, Y., Abbeel P., 2017. *Adversarial Attacks on Neural Network Policies.* Computing Research Repository. Available here. Behzadan, V., Munir, A., 2017. *Vulnerability of Deep Reinforcement Learning to Policy Induction Attacks*. Computing Research Repository. Available here.

39 See for all OpenAI's excellent description of the problem. Goodfellow, I., Papernot, N., Huang, S., Duan, Y., Abbeel, P., Clark, J., 2017. *Attacking Machine Learning with Adversarial Examples*. OpenAI Blog. Available here.

40 Goodfellow, I., Papernot, N., Huang, S., Duan, Y., Abbeel, P., Clark, J., 2017. *Attacking Machine Learning with Adversarial Examples*. OpenAI Blog. Available here. Papernot, N. & Goodfellow, I., 2016. *Breaking things is easy*. Cleverhans-Blog. Available here. Goodfellow, I., & Papernot, N., 2017. *Is attacking machine learning easier than defending it?* Cleverhans-Blog. Available here.

41 Goodfellow, I., & Papernot, N., 2017. *Is attacking machine learning easier than defending it?* Cleverhans-Blog. Available here. Goodfellow, I., Shlens, J., Szegedy, C., 2015. *Explaining and Harnessing Adversarial Examples*. International Conference on Learning Representations. Available here. Goodfellow, I., Papernot, N., Huang, S., Duan, Y., Abbeel, P., Clark, J., 2017. *Attacking Machine Learning with Adversarial Examples*. OpenAI Blog. Available here.

autonomous stockpiling of software vulnerabilities (e.g. zero-day attacks) and the proliferation of malicious code into the global Internet in ways never envisioned (or subsequently controllable) by human designers (see, e.g. Stuxnet). AI-enhanced cyber-attacks will have an asymmetrical advantage over exclusively human operators.[42] Using machine learning approaches, these systems will automatically decide on the most effective attack and defense vectors. For diplomats engaged for the last few years in an effort to establish norms of cyber law, to gain cooperation in the investigation and prosecution of cyber-crime, and to define unlawful cyber-attacks that are subject to sanction by the international community – these tasks will become far more challenging.

In addition to expanded arenas of cyber-attack, there will likely be a broader set of information operations that aim to deceive, disrupt, and distort public communications in enemy states. The leading edge of information operations is already visible in the alleged Russian operation to influence the 2016 US election by leveraging the power of AI-enhanced social media ad targeting and armies of automated accounts on Facebook and Twitter. On one level, this represents a technology-driven escalation of the age-old practice of propaganda. But the effectiveness of AI-empowered techniques have led to a very significant reevalaution of election security and the integrity of the public debate in democracies. In the future, "data warfare" may include a virtual battle between artificial intelligences seeking to disable one another and infect command and control systems with disinformation or malicious code. It may include sophisticated media forgeries[43] developed through AI designed to dupe the opposing public into relying on falsehoods or acting contrary to their interests. Damage to the integrity of democratic discourse and the reputation of state institutions and their representatives will be easier to inflict and harder to repair. Diplomats will be tasked with responding to fallout from all of these challenges.

Preliminary Policy Agenda

The top priority in this area is updating arms control and non-proliferation strategies to deal with an escalating AI arms race. In particular, this means aligning major powers around common policies (such as limitations on

42 Allen, G. & Chan, T., 2017. *Artificial Intelligence and National Security*. Belfer Center for Science and International Affairs Harvard Kennedy School. Cambridge. Available here.

43 ibid.

offensive capabilities) and working together in the common interest of guarding against these weapons falling into the hands of terrorists. This work should be accompanied by significant public diplomacy to establish moral red lines and convene influential stakeholders across sectors to contain the threat of AI weapons. In addition, there is much work to be done evaluating the potential threats of AI in hard power as well as in disinformation campaigns. There is too little understanding in our ministries about how these technologies work, which players in which markets offer weaponized AI as a product, and how we might be able to push back against them.

1. Drawing red lines. Governments have begun planning and investing for the AI future, but none have yet developed and articulated red lines about how AI technologies may be used according to the norms of international law and human rights. This should be a national as well as international effort.
2. Public diplomacy on AI Ethics. There is a window of opportunity to begin coordinated efforts at global public communications to raise awareness about the ethics of autonomous killing to establish norms across government, industry, academia, and civil society organizations focused on these issues.
3. Adapting arms control. The foreign policy community should immediately intensify nascent efforts to develop a new regime of multilateral arms control. This requires a new set of considerations for export controls, dual-use criteria, and enforcement measures. This work must not only be government-to-government, but also government-to-business and engaging the AI R&D community to help design these safeguards into future products.
4. Combating Disinformation. To address the growing problems of disinformation and divisive propaganda, the foreign policy community could lead an international discussion about developing standards of trustworthy communications among states and peoples. This dialogue would be designed to open spaces for public communications that are secure from exploitation and grant a channel to dispel conspiracy and disinformation. Further, it would offer information sharing about the AI-powered tools of digital deception and best practices to counter them.

Topic #3 – Democracy and Ethics

Strategic Priorities

The job of foreign ministries in most liberal democracies includes two straightforward and related tasks that reflect the values of open societies. The first is to promote and strengthen democratic institutions that protect social equality and representation around the world. The second is to pursue a (human and civil) rights-based system of governance, commerce, and security in the international community. The emergence of AI technologies poses both serious challenges and inspiring new opportunities to both of these objectives. Once again, we see the priority areas for foreign policy planners as focused on two sets of risk mitigation issues: 1) AI-enhanced surveillance practices that may constrain civil rights and liberties; 2) socio-cultural conflict that may be deepened through the perpetuation of social bias and discrimination rooted in AI algorithms.

AI Restrictions on Rights and Liberties

The relationships between AI technologies and civil liberties is an area fraught with tension for diplomats. They are charged both with advancing security interests that privilege control technologies and a human rights agenda that seeks to enhance freedom through technology. The Internet poses an analogous dilemma and has clear parallels worthy of study. One obvious example of this challenge came with the diplomatic fallout from the Snowden revelations of global surveillance by Western intelligence agencies that drew such a sharp contrast to the same governments' work to promote Internet freedom. This time around, there is no way to downplay or evade the potential of using AI to enhance surveillance. The tension between security and liberty must be reconciled in a common strategic plan.

The AI transformation in data processing – including facial and voice recognition at scale, code breaking, and fact-pattern correlation – is a game-changer for intelligence and law-enforcement surveillance operations. For liberal democracies, this raises a set of ethical questions about the constraints placed on this power and the establishment of meaningful oversight. Clearly, it will not be the task of foreign policy to design and implement checks and balances on the surveillance practices of security agencies. But it will fall to international diplomacy to communicate these

policies to the world in pursuit of a moral credibility that can support leadership on a human rights agenda.

In non-democratic states, the near-term impact is more ominous. With its interest in surveillance and censorship driven by concerns for national security, China has emerged as a leader in AI-enabled surveillance. Chinese governmental interests in this field are also animated by commercial investments. Among industry leaders in facial recognition software are top Chinese firms such as Baidu[44], Tencent, and Sensetime. These companies can train their algorithms on vast amounts of user generated data (the country has more than 700 million internet users).[45] Face recognition software from China excels in international competitions on the accuracy of these AI-enabled systems. Yitu Tech, a Chinese startup, is the latest example for very accurate face recognition performance under difficult test conditions. Yitu has recently won the Face Recognition Prize Challenge[46], hosted by the Intelligence Advanced Research Projects Activity (IARPA) under the U.S. Office of the Director of National Intelligence (ODNI).[47] On maturity, this technology may be purchased by law enforcement agencies, allowing them to cross-reference the images of social media with a centralized image database of citizens.[48] In addition to user-generated data, Megvii, a Chinese startup that is specialised in facial recognition, is training its machine learning algorithms on data of facial scans drawn from a database of the Ministry of Public Security.[49] This database holds facial data on 1.3 billion Chinese citizens.[50] This could inspire other authoritarian states to follow China's

44 Over the last 2.5 years, Baidu has invested $1.5 billion in AI research ("in addition to $200 million it committed to a new in-house venture capital fund, Baidu Venture"). See McKinsey Global Institute, 2017. *Artificial Intelligence. The next Digital Frontier?* McKinsey Global Institute. Available here.

45 Chin, J. & Lin, L., 2017. *China's All-Seeing Surveillance State Is Reading Its Citizens' Faces*. The Wall Street Journal. Available here.

46 Challenge.gov, 2017. Face Recognition Prize Challenge. U.S. General Services Administration. Available here.

47 CISION PR Newswire, 2017. Yitu Tech Wins the 1st Place in Identification Accuracy In Face Recognition Prize Challenge 2017. PR Newswire. Available here.

48 All Chinese citizens are required by law to carry a government-issued photo ID as early as the age of 16. Chin, J. & Lin, L., 2017. *China's All-Seeing Surveillance State Is Reading Its Citizens' Faces*. The Wall Street Journal. Available here.

49 Chen, L. Y., 2017. *China, Russia Put Millions in This Startup to Recognize Your Face.* BloombergTechnology. Available here.

50 ibid.

lead and to control their people with AI enhanced surveillance systems. This is likely to go far beyond tracking tools and privacy rights. AI is already an influential part of censorship regimes that seek to identify and delete online content that is unwanted by the government.

But we also see growing interest in these technologies by law enforcement and national security agencies around the rest of the world. The proliferation of surveillance-focused AI technologies is likely to increase the frequency of episodes such as the recent scandal in Mexico that uncovered government surveillance of journalists using Israeli-made malware.[51] For foreign policy planners, these developments signal an intensification of the "Snowden contradictions" and an increasing need to protect the privacy and communication rights of journalists, dissidents, and civil society activists in illiberal states.

AI Bias and Discrimination

In our effort to support economic growth and prosperity through the growth of AI technology markets, products and services will proliferate that have unknown and untested social consequences. There is considerable research currently focused on the unintended consequences of automated decision-making tools that may replicate and deepen existing social discrimination.[52] We must seek to anticipate these problem, raise awareness, and promote a measure of social equity in technical design. AI systems derive their intelligence from supervised or unsupervised learning experiences. They infer their logic on the basis of vast amounts of ingested training data and the original parameters used for the development of the algorithms. Consequently, their reasoning and their actions reflect the quality of the data that was used for training as well as the biases of the programmers. Often the scope and depth of the data is inadequate to reflect the complexity of

51 Deibert, R., 2017. Mexico Wages Cyber Warfare Against Journalists and their minor children. *Ronald Deibert*. Available here.

52 See, for example Crawford, K. & Whittaker, M., 2016. *The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*. AI Now, New York University. Available here.

problems. These shortcomings become truly problematic when those who develop and use AI are not aware of these limitations and biases.[53]

As we increasingly use AI technologies to make important decisions about access and allocation of social and economic equities – for example credit-worthiness or the evaluation of job applications – we risk undermining basic values of fairness and non-discrimination in both the private and public sector.[54] Although, AI has the potential to increase fairness in decision-making by removing some elements of human bias, there is a converse logic of injecting the bias built into the machine. The recent debates about predictive policing[55] are exemplary for larger concerns that AI could undermine civil rights by reproducing biases relative to race, ethnicity, gender, sexual orientation, and income. Foreign Ministries should evaluate how these debates reflect important risks to democratic institutions and civil rights and raise these issues with other governments. In their own practice, ministries that embrace data-driven AI tools for development aid projects (a likely, and potentially fruitful, prospect for the medium term) should keep the problem of bias front of mind.

Preliminary Policy Agenda

The diplomatic and development agenda surrounding the Internet has demonstrated for years the tensions between security and freedom implicit in ever more connected societies. AI will heighten this tension by supercharging surveillance and censorship capabilities. Even as these technologies enable new opportunities for free expression, civic activity, and social progress, they also raise the unwelcome possibility of deepening existing social discrimination. The challenge for foreign policy will be to promote a positive agenda in the face of these risks – leveraging grant-making, communications, and multi-lateral policy engagement to pursue rights-based goals.

1. Grantmaking in AI and Human/Civil Rights: Much as foreign policy has promoted Internet freedom through grant-making that enables and spreads technologies that support secure and private communications, we should consider new programs that allocate resources to research organizations around the world that can audit and measure the impact of

53 For reasons of "inaccurate measurement methodologies, incomplete data gathering, non-standardized self-reporting, or other flaws in data collection", for details see: ibid.

54 ibid.

55 Joh, E. E., 2017. Feeding the Machine: Policing, Crime Data, & Algorithms. William & Mary Bill of Rights J. (2017 Forthcoming). Available here.

AI technologies coming into global markets. This should include a strong measure of both the spirit and practice of Internet freedom policies in the new climate of AI-led surveillance.

2. Public Diplomacy: Ministries should also leverage their public diplomacy tools to raise public awareness about both the benefits and the risks of AI in our societies. An ethical AI communications strategy not only fosters soft power around these technologies, it is a way to positively characterize and differentiate domestic AI products and services in a world of governments and peoples that may grow wary of the opaque power of AI's leading corporations.
3. Rights-Based Data Policy: In the context of regional and global economic policy as well as trade negotiations, foreign ministries will have a clear opportunity to raise issues of AI and social discrimination. Much as intellectual property, cyber-security, and data privacy have become elements of global economic dialogue, a rights-based AI agenda should be integrated into these institutions and processes.

Conclusion

At present, the focus of scholarship, investment and political debate about AI is on markets and weapons. In a welcome turn, there is also a growing sector of research and advocacy with respect to AI, public policy, and basic ethical question about how societies should evaluate and manage the consequences of a world driven by automated decisions. By contrast, research and analysis at the intersection of AI and foreign policy is quite underdeveloped. Commentary on the role of diplomacy and statecraft is scarce. And yet, clearly there are major implications for policy development and programmatic work for the foreign service.

In this paper, we offered an outline of the foreign policy challenges implicit in three areas where AI will have a powerful impact: global economics, international security, and democratic ethics. We chose these because they are traditionally core areas of work for foreign ministries. We set a brief analysis of these issues on top of guidelines for how to ground foreign policy making and diplomatic practice focused on the impacts of AI on international relations. To serve the ends of policy planners, we concluded each section with a preliminary policy agenda to provide starting points for future work in this area.

Grand theory about technology-driven change at the global level must be instrumented through institutions. And we recognize that these institutions

operate under constraints – political, budgetary, bureaucratic, and human resources. Consequently, we opted to present a pragmatic proposal for the foreign policy of AI that leverages the existing tools of diplomacy while working towards more systemic adaptation in the future. Although we believe that transformational changes to our diplomatic institutions will eventually be needed to meet the challenges ahead, we see the best path forward as an incremental approach to AI that builds on the successes (and learns from the failures) of "cyber-foreign policy". In most countries, this work on cyber issues is now operationalized and there is a base of familiarity within the institution from which planners can work on the next technology revolution. This should be a holistic effort to address the role of technology across governmental responsibilities and ministerial equities. It is a policy planning process, a programmatic development and implementation strategy, and an HR challenge to sustain this work over time. This work on the statecraft of the Internet age is a significant achievement in a relatively short period of time. We must now do the same for AI, but we cannot afford to spend a decade thinking about it.

Workshop participants

Gregory C. Allen	Center for a New American Security
Ralf Beste	Federal Foreign Office
Damian Borth	German Research Center for Artificial Intelligence
Kate Crawford	AI Now Institute / New York University / Microsoft Research
Axel Gugel	Federal Foreign Office
Stefan Heumann	Stiftung Neue Verantwortung
Mirko Hohmann	Global Public Policy Institute
Kirsten Hommelhoff	Stiftung Mercator
Tilo Klinner	Federal Foreign Office
Philippe Lorenz	Stiftung Neue Verantwortung
Marcel A. Mayr	Futurist / Technology and AI Researcher
Trent McConaghy	Ocean Protocol / BigchainDB GmbH
Heiko Nitzschke	Federal Foreign Office
Christoph Peylo	Robert Bosch GmbH
Michael Schwarz	Stiftung Mercator
Ben Scott	Stiftung Neue Verantwortung
Katharina Semmler	Stiftung Mercator
Ludwig Siegele	The Economist
Matthias Spielkamp	Algorithm Watch
Fabian J.G. Westerheide	Asgard Capital / German Startup Association
Meredith Whittaker	AI Now Institute / New York University / Google Open Research

Recommended Reading on AI & Foreign Policy

Economics

Acemoglu, D. & Restrepo, P., 2017. Robots and jobs: Evidence from the US. VOX, CEPR's Policy Portal. Available at: http://voxeu.org/article/robots-and-jobs-evidence-us

Agrawal, A., Gans, J. & Goldfarb, A., 2016. The Simple Economics of Machine Intelligence. Harvard Business Review. Available at: https://hbr.org/2016/11/the-simple-economics-of-machine-intelligence

Autor, D.H. & Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the US Labor Market. American Economic Review, 103(5), pp.1553–1597. Available at: http://www.ddorn.net/papers/Autor-Dorn-LowSkillServices-Polarization.pdf.

Avent, R., 2016. The Wealth of Humans: the Future of Work in the Twenty-first Century. St Martins Pr.

Brown, J. & Lorenz, P., 2017. The Future of Work and the Trans-Atlantic Alliance. Bertelsmann Foundation North America, Washington, D.C. and Stiftung Neue Verantwortung, Berlin. Available at: https://www.stiftung-nv.de/sites/default/files/the_future_of_work_the_trans.pdf.

Brynjolfsson, E. Rock, D., Syverson, C., 2017. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. National Bureau of Economic Research. Available at: https://www.nber.org/papers/w24001.

Dauth, W., Findeisen S., Südekum J., Woessner N., 2017. German Robots - The Impact of Industrial Robots on Workers. Centre for Economic Policy Research. Available at: cepr.org/active/publications/discussion_papers/dp.php?dpno=12306.

Dauth, W., Findeisen S., Südekum J., Woessner N., 2017. The rise of robots in the German labour market. VOX, CEPR's Policy Portal. Available at: http://voxeu.org/article/rise-robots-german-labour-market.

Executive Office of the President, 2016. Artificial Intelligence, Automation, and the Economy. The White House. Available at: https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Executive Office of the President, National Science and Technology Council, 2016. Preparing for the Future of Artificial Intelligence. The White House. Available at: https://obamawhitehouse.archives.gov/sites/default/files/

whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
Knight, W., 2017. China's AI Awakening 中国 人工智能 的崛起. MIT Technology Review. Available at: https://www.technologyreview.com/s/609038/chinas-ai-awakening/?utm_campaign=add_this&utm_source=twitter&utm_medium=post
Lorenz, P., 2017. Digitalisierung im deutschen Arbeitsmarkt - Eine Debattenübersicht. Stiftung Neue Verantwortung. Berlin. Available at: https://www.stiftung-nv.de/sites/default/files/snv_digitalisierung_arbeitsmarkt_philippe_lorenz_langversion.pdf
McKinsey Global Institute, 2017. A Future that Works: Automation, Employment, and Productivity. McKinsey Global Institute. Available at: https://www.mckinsey.com/~/media/McKinsey/Global%20Themes/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works_Full-report.ashx.
McKinsey Global Institute, 2017. Artificial Intelligence. The next Digital Frontier? McKinsey Global Institute. Available at: https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx.
Metz, C., 2017. Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent. The New York Times. Available at: https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html?_r=0.
Morikawa, M., 2016. Artificial intelligence and employment. VOX, CEPR's Policy Portal. Available at: http://voxeu.org/article/artificial-intelligence-and-employment.

Security

Allen, G. & Chan, T., 2017. Artificial Intelligence and National Security. Belfer Center for Science and International Affairs Harvard Kennedy School. Cambridge. Available at: https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf.
Biegel, B. & Kurose, J.F., 2016. The National Artificial Intelligence Research and Development Strategic Plan. The White House. Available at: https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf.
Bostrom, N. 2013. Superintelligence: Paths, Dangers, Strategies. Oxford, Oxford University Press.
De Spiegeleire, S., Maas, M., Sweijs, T., 2017. Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers. The Hague Centre for Strategic Studies (HCSS). Available

at: http://hcss.nl/sites/default/files/files/reports/Artificial%20Intelligence%20and%20the%20Future%20of%20Defense.pdf.
Geiss, R., 2015. The International-Law Dimension of Autonomous Weapons Systems. Friedrich-Ebert-Stiftung. Berlin. Available at: http://library.fes.de/pdf-files/id/ipa/11673.pdf.
Jason, 2017. Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD. The MITRE Corporation. Virginia. Available at: https://fas.org/irp/agency/dod/jason/ai-dod.pdf.
Scharre, P. & Horowitz, M.C., 2015. An Introduction to Autonomy in Weapons Systems. Center for a New American Security. Available at: https://s3.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf?mtime=20160906082257.
Schmitt, M.N. & Thurnher, J.S., 2013. "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict. Harvard National Security Journal, 4(2), pp.231–281. Available at: http://harvardnsj.org/wp-content/uploads/2013/01/Vol-4-Schmitt-Thurnher.pdf.
Schmitt, M.N. & Thurnher, J.S., 2013. "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict. Harvard National Security Journal, 4(2), pp.231–281. Available at: http://harvardnsj.org/wp-content/uploads/2013/01/Vol-4-Schmitt-Thurnher.pdf.
Simonite, T., 2017. AI could revolutionize War as much as Nukes. Wired. Available at: https://www.wired.com/story/ai-could-revolutionize-war-as-much-as-nukes/.
The Economist, 2017. Code red. Why China's AI push is worrying. The Economist. Available at: https://www.economist.com/news/leaders/21725561-state-controlled-corporations-are-developing-powerful-artificial-intelligence-why-chinas-ai-push.
Webster, G., Creemers, R., Triolo, P., Kania, E., 2017. China's Plan to 'Lead' in AI: Purpose, Prospects, and Problems. New America. Available at:https://www.newamerica.org/cybersecurity-initiative/blog/chinas-plan-lead-ai-purpose-prospects-and-problems/.
Webster, G., Creemers, R., Triolo, P., Kania, E., 2017. State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan. Available at: https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17.pdf.

Ethics

Anderson, K., Waxman, M. & Perkins, J., 2013. Law and Ethics for Autonomous Weapon Systems. Available at: http://media.hoover.org/sites/default/files/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf.

Babcock, J., Kramar, J. & Yampolskiy, R. V., 2017. Guidelines for Artificial Intelligence Containment. Available at: https://arxiv.org/abs/1707.08476.

Beard, J.M., 2014. Autonomous Weapons and Human Responsibilities. Available at: https://www.law.georgetown.edu/academics/law-journals/gjil/recent/upload/zsx00314000617.PDF.

Chin, J. & Lin, L., 2017. China's All-Seeing Surveillance State Is Reading Its Citizens' Faces. The Wall Street Journal. Available at: https://www.wsj.com/articles/the-all-seeing-surveillance-state-feared-in-the-west-is-a-reality-in-china-1498493020.

Crawford, K. & Whittaker, M., 2016. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term. AI Now, New York University. Available at: https://assets.contentful.com/8wprhhvnpfc0/3JOy5k4f1YSCQOi8MCCmA2/97010d04fbc7892662ce8b2469dc1601/AI_Now_2016_Report.pdf.

European Parliament. Committee on Legal Affairs, 2016. Motion for A European Parliament Resolution with Recommendations to the Commission On Civil Law Rules on Robotics, 2015/2103(INL). Available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN .

Knight, W., 2017. The Dark Secret at the Heart of AI. MIT Technology Review. Available at: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

Ng, A., 2017. What Artificial Intelligence can and can't do right now. Harvard Business Review. Available at: https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now.

Tegmark, M., 2017. Life 3.0 being human in the age of artificial intelligence. New York, Alfred A. Knopf.

Various, 2017. An Open Letter to the United Nations Convention on Certain Conventional Weapons. Available at: https://www.cse.unsw.edu.au/~tw/ciair/open.pdf.

About Stiftung Neue Verantwortung

The Stiftung Neue Verantwortung (SNV) is an independent think tank that develops concrete ideas as to how German politics can shape technological change in society, the economy and the state. In order to guarantee the independence of its work, the organisation adopted a concept of mixed funding sources that include foundations, public funds and businesses.

Issues of digital infrastructure, the changing pattern of employment, IT security or internet surveillance now affect key areas of economic and social policy, domestic security or the protection of the fundamental rights of individuals. The experts of the SNV formulate analyses, develop policy proposals and organise conferences that address these issues and further subject areas.

About the Authors

Dr. Ben Scott

Ben Scott is a member of the management board of the Stiftung Neue Verantwortung. He is also Senior Adviser at New America's Open Technology Institute in Washington DC. During the 2016 presidential election in the US, he served as the coordinator of technology and innovation policy advisers for the Hillary Clinton campaign. From 2010-2012, Ben Scott was Policy Advisor for Innovation at the US Department of State where he worked at the intersection of technology and foreign policy to steward Secretary Clinton's 21st Century Statecraft agenda. Prior to joining the State Department, for six years he led the Washington office for Free Press, a public interest organization in the US dedicated to protecting the open Internet and public service journalism. Before joining Free Press, he worked as a legislative aide handling telecommunications policy for then-Rep. Bernie Sanders in the U.S. House of Representatives. He holds a PhD in communications from the University of Illinois.

bscott@stiftung-nv.de
+49 (0)30 81 45 03 78 80

Dr. Stefan Heumann

Stefan Heumann is member of the management board of Stiftung Neue Verantwortung (SNV). Prior to joining the board, he initiated and built the European Digital Agenda program together with Ben Scott and directed it until March 2016, laying the foundation for the further strategic development of the SNV into a think tank working at the intersection of technology and society. He is a member of the advisory board of technology policy assessment of the German National Academy of Science and Engineering (acatech). From 2014 to 2016 he was a member of the Freedom Online Coalition's working group 3 on „privacy and transparency online". Before joining SNV in 2013, he coordinated the public affairs section of the US Consulate General in Hamburg. From 2009 to 2010 he taught and researched political science as Assistant Professor at the University of Northern Colorado. Stefan holds a PhD from the University of Pennsylvania and studied political science at the Free University of Berlin, the University of the Provence in Aix-en-Provence, and the University of Pennsylvania in Philadelphia.

sheumann@stiftung-nv.de
+49 (0)30 81 45 03 78 80

Philippe Lorenz

Philippe Lorenz analyzes the evolution of the German labor market under the influence of rapid technological change. He was part of the foresight lab 'Toward the Labor Market 4.0?', a joint research project between the Stiftung Neue Verantwortung and the Bertelsmann Stiftung, which assessed the potential impacts of digitization on labor and employment in Germany until 2030. In addition to his work on labor politics, he examines the impact of digital technologies on the German energy revolution (Energiewende). Philippe studied law at the University of Passau and International Relations at the Rhine-Waal University of Applied Sciences in Kleve.

plorenz@stiftung-nv.de
+49 (0)30 81 45 03 78 94

Imprint

stiftung neue verantwortung e. V.

Beisheim Center
Berliner Freiheit 2
10785 Berlin

T: +49 (0) 30 81 45 03 78 80
F: +49 (0) 30 81 45 03 78 97

www.stiftung-nv.de
info@stiftung-nv.de

Design:
Make Studio
www.make-studio.net

Layout:
Johanna Famulok

Kostenloser Download:

www.stiftung-nv.de