# 51. [The use of generative AI and AI-assisted technologies in writing for Elsevier](https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier)

# The use of generative AI and AI-assisted technologies in writing for Elsevier

Last updated 18 August 2023

## Policy for Book and Commissioned Content Authors

This policy aims to provide greater transparency and guidance to authors, readers, reviewers, editors in relation to generative AI and AI-assisted technologies.  Elsevier  will monitor this development and will adjust or refine this policy when appropriate. Please note the policy only refers to the writing process, and not to the use of AI tools to analyze and draw insights from data as part of the research process.

Where authors use AI and AI-assisted technologies in the writing process, these technologies should only be used to improve readability and language of the work and not to replace key authoring tasks such as producing scientific, pedagogic, or medical insights, drawing scientific conclusions, or providing clinical recommendations. Applying the technology should be done with human oversight and control and all work should be reviewed and edited carefully, because AI can generate authoritative-sounding output that can be incorrect, incomplete, or biased. The authors are ultimately responsible and accountable for the contents of the work.

Authors should disclose in their manuscript the use of AI and AI-assisted technologies and a statement will appear in the published work. Declaring the use of these technologies supports transparency and trust between authors, readers, reviewers, editors, and contributors and facilitates compliance with the terms of use of the relevant tool or technology.

Authors should not list AI and AI-assisted technologies as an author or co-author, nor cite AI as an author. Authorship implies responsibilities and tasks that can only be attributed to and performed by humans. Each (co-) author is accountable for ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved and authorship requires the ability to approve the final version of the work and agree to its submission. Authors are also responsible for ensuring that the work is original, that the stated authors qualify for authorship, and the work does not infringe third party rights, and should familiarize themselves with [Elsevier’s Publishing Ethics policy](https://www.elsevier.com/about/policies-and-standards/publishing-ethics-books) before they submit.

## The use of generative AI and AI-assisted tools in figures, images and artwork

Elsevier does not permit the use of generative AI or AI-assisted tools to create or alter images in submitted manuscripts. This may include enhancing, obscuring, moving, removing, or introducing a specific feature within an image or figure. Adjustments of brightness, contrast, or color balance are acceptable if they do not obscure or eliminate any information present in the original. Image forensics tools or specialized software might be applied to submitted manuscripts to identify suspected image irregularities.

The only exception is if the use of AI or AI-assisted tools is part of the research design or research methods (such as in AI-assisted imaging approaches to generate or interpret the underlying research data, for example in the field of biomedical imaging). If this is done, such use must be described in a reproducible manner in the methods section. This should include an explanation of how the AI or AI-assisted tools were used in the image creation or alteration process, and the name of the model or tool, version and extension numbers, and manufacturer. Authors should adhere to the AI software’s specific usage policies and ensure correct content attribution. Where applicable, authors could be asked to provide pre-AI-adjusted versions of images and/or the composite raw images used to create the final submitted versions, for editorial assessment.

The use of generative AI or AI-assisted tools in the production of artwork such as for book or commissioned content covers or graphical abstracts is not permitted.

# 52. [AI workplace changes](https://www.nytimes.com/2023/05/08/business/dealbook/ai-workplace-changes.html)

By midyear, all of Morgan Stanley’s thousands of wealth advisers are expected to have access to a new artificial-intelligence-powered chat tool.

The tool, which is already in use by about 600 staff members, gives advisers answers to questions such as “Can you compare the investment cases for Apple, IBM and Microsoft?” and follow-ups such as “What are the risks of each of them?” An adviser can ask what to do if a client has a potentially valuable painting — and the knowledge tool might provide a list of steps to follow, along with the name of an internal expert who can help.

“What we’re trying to do is make every client or every financial adviser as smart as the most knowledgeable expert on any given topic in real time,” said Jeff McMillan, the head of analytics, data and innovation for Morgan Stanley Wealth Management.

Experts disagree about whether A.I. will wind up destroying more jobs than it creates over time. But it is clear that A.I. will alter work for most knowledge workers, shifting the skills they need and changing the staffing needs of most companies. Now it’s up to business leaders to figure out how to take advantage of the technologies today, while preparing workers for the disruption that the tools present over the medium term.

Moving too slowly may mean losing out on gains in productivity, customer service and — ultimately — competitiveness, similar to what happened to businesses that didn’t embrace the internet fully or fast enough. But at the same time, leaders mustguard against the mistakes and biases A.I. often perpetuates and be thoughtful about what it means for employees.

“Almost no matter which sector you are in, you need to be thinking about your company as becoming an A.I.-first company,” said Alexandra Mousavizadeh, chief executive at [Evident](https://evidentinsights.com/), a start-up that analyzes finance companies’ A.I. capabilities.

# 53. [Guidelines establish proper uses of AI in research](https://www.chinadaily.com.cn/a/202401/05/WS65981e57a3105f21a507ad78.html)

The Ministry of Science and Technology recently issued guidelines on the proper use of artificial intelligence-related content, emphasizing that researchers should integrate ethical requirements into the entire research process.

The Guidelines for Responsible Research Conduct (2023) state that generative AI should not be used to directly generate research application materials, nor should it be listed as a co-author.

They also state that when using AI to generate content, especially when it involves facts and opinions, it should be clearly labeled and the generation process should be explained to ensure accuracy, authenticity and respect for intellectual property rights.

In addition, content generated by AI that has been labeled as such by other authors should generally not be cited as original references. Unverified references generated by AI should not be used directly. The primary methods and details of using AI should be disclosed in the research methods or appendix sections.

During the publication of research outcomes, authors should disclose whether generative AI was used, specify the software name, version and usage time, and provide specific annotations for auxiliary generated content related to facts and opinions, the guideline says.

In the peer review process, if generative AI is to be used, prior consent should be obtained from the organizers of the review activity. Measures should be taken to prevent the leak of review content, and necessary remedial actions should be taken promptly in case of an information leak, according to the guidelines.

The guideline applies to research institutions, universities, medical and health institutions, enterprises, and their researchers. The specified requirements cover the main aspects and processes of scientific and technological activities.

The ministry stated that conducting responsible scientific research is a widely recognized standard in international science and technology governance, and is an essential requirement for promoting the healthy development of China's science and technology industry and achieving high-level technological self-reliance.

# 54. [China unveils new artificial intelligence guidelines for scientists and bans use in funding applications](https://www.scmp.com/news/china/science/article/3247420/china-unveils-new-artificial-intelligence-guidelines-scientists-and-bans-use-funding-applications)

* The guidelines say AI cannot be credited as a co-author and any use of the technology for research purposes must be clearly labelled
* Some scientists welcomed the move, with one saying he feared research students could be ‘ruined’ by misusing the technology
* China has published new guidelines on the use of generative [artificial intelligence](https://www.scmp.com/topics/artificial-intelligence?module=inline&pgtype=article) in scientific research, including a ban on the “direct” use of the technology when applying for research funding and approval.
* The Guidelines For Responsible Research Conduct were released by the Ministry of Science and Technology on December 21, with the aim of promoting the “healthy development” of scientific research.
* The guidelines cover various aspects of the research process, including topic selection and peer review, with ethics, safety and transparency the main considerations.
* The rules apply to individual researchers and institutions, including universities and medical and health centres.
* The guidelines set out a general framework for the use of AI, but do not provide detailed rules for specific situations.
* One of the main concerns is the use of AI-generated content, which cannot be listed as a co-author under the new rules.
* AI has helped researchers around the world make discoveries such as [new antibiotics that may help fight drug-resistant superbugs](https://www.scmp.com/news/world/united-states-canada/article/3247200/potent-antibiotic-has-emerged-battle-against-deadly-drug-resistant-superbugs?module=inline&pgtype=article) and a [Chinese team’s AI-generated catalysts that could help produce oxygen on Mars](https://www.scmp.com/news/china/science/article/3241505/chinese-scientists-say-ai-chemist-can-create-catalysts-make-oxygen-martian-meteorites?module=inline&pgtype=article).
* Some scientists have also listed AI tools such as ChatGPT as co-authors – a practice many journals have already stopped.
* However, this has prompted discussions about whether AI should be credited if it discovers new materials or drugs – and even a broader philosophical debate about whether it should be considered equal to humans in the pursuit of knowledge.
* Under the guidelines, generative AI can still be used in research, but any content or findings that use the technology must be clearly labelled as such.
* Wen Shaoqing, an associate professor at Fudan University in Shanghai, said the effect of the regulations was likely to be “limited”, at least in his field of scientific archaeology, because generative AI “tends to produce poor-quality project application proposals” and few people will use it.
* A biologist from the Chinese Academy of Sciences, who asked not to be named, also said she never uses AI to help her write proposals because these language models produce “beautiful but essentially meaningless words”.
* However, another researcher from the Chinese Academy of Social Sciences said the widespread use of generative AI “is an irreversible trend” and it is unrealistic and unreasonable to ban it across the board.
* The scientist, who also spoke on condition of anonymity, said researchers will continue to use it in research and grant applications if they think it will help them and it is more important to figure out “how to allow researchers to use it in an appropriate way and to … keep up with international trends”.
* The rapid development of AI technology has “promoted profound changes in scientific research paradigms”, according to a document from the science ministry explaining the new rules.
* However, it continued that the guidelines were needed because the technology has created problems around authorship, intellectual property and data processing.
* According to the ministry, the guidelines draw on “useful foreign experiences and reflect international practices.”
* It said the guidelines were based on a “broad consensus” in scientific circles, and will be updated and adjusted based on further technological developments.
* Regulating the use of AI became a major focus for Beijing last year, both domestically and internationally.
* In April a meeting of the Politburo said it attached “great importance” to developing AI but highlighted the importance of risk prevention.
* In August [the country’s first regulations](https://www.scmp.com/tech/big-tech/article/3227576/china-sets-out-new-rules-generative-ai-beijing-emphasising-healthy-content-and-adherence-socialist?module=inline&pgtype=article) – the Interim Measures for the Management of Generative AI Services – came into effect with the aim of promoting “socialist values” and the “healthy” development of publicly available AI content.
* These measures do not apply to scientific research institutes as long as they do not provide generative AI services to the Chinese public, according to the Cyberspace Administration of China.
* At the start of November, China along with 27 other countries, including the United States, and the European Union, met in Britain to sign the [Bletchley Declaration, an international pact on regulating the potentially “catastrophic” risks](https://www.scmp.com/news/china/diplomacy/article/3240306/ai-deal-shows-china-and-us-can-work-together-tech-regulation-despite-rivalry-analysts-say?module=inline&pgtype=article) that could arise from rapidly developing AI.
* China’s new scientific guidelines expand upon the interim measures. Apart from the ban on the direct use of AI when writing applications, the rules also say that AI-generated content and results must be labelled within the text “especially when it involves key content such as facts and opinions”.
* AI-generated content should also be identified in the footnotes, methods section or appendices of research papers, along with explanations of how it was created and what software was used.
* Any content that is marked as being generated by AI should not be treated as original literature, and if other authors want to cite this content “an explanation should be given”, the guidelines say.
* They also state that any references generated by AI cannot be used unless they are verified first. A library guide from University College London has previously warned that generative AI may fabricate references and quotes.
* A scientist from the Chinese Academy of Sciences working in organic chemistry said restrictions on AI-generated content are necessary because many graduate students now use it to write their theses, and he fears “it will ruin these students”.
* The scientist, who asked not to be named, believes that such tools can make students lazier and makes it harder for them to develop original thinking skills.

# 55. [Penguin Random House CEO hopes AI will help sell more books: Report](https://www.thehindu.com/sci-tech/technology/penguin-random-house-ceo-hopes-ai-will-help-sell-more-books-report/article67796207.ece)

Penguin Random House CEO Nihar Malaviya said that he hoped AI would make it easier to sell more book titles without hiring more people in the future, reported The New York Times this week. The report was a profile of the chief of Penguin Random House - one of the ‘Big Five’ global publishing companies that employs thousands of professionals and dominates the English-language mainstream publishing sector. The piece noted that Malaviya was focusing on growth, with the company laying off around 60 people last year as well as using voluntary buyouts. enguin Random House tried to acquire fellow Big Five publisher Simon & Schuster in 2020, but was blocked on antitrust grounds, leading to the deal being dropped around two years later.

*(For top technology news of the day,*[*subscribe*](https://www.thehindu.com/newsletter-subscription/?nl_id=4307&utm_source=website_article_todays_cache&utm_campaign=newsletter_signup&utm_medium=website)*to our tech newsletter Today’s Cache)*

Publishers in India are also looking at artificial intelligence to help book production. Harper Collins India CEO Ananth Padmanabhan noted last year that AI narrators could bring about quicker releases of translated novels.

While [publishers worldwide explore the use of large language models](https://www.thehindu.com/sci-tech/technology/apple-explores-ai-deals-with-news-publishers-report/article67668620.ece) and the AI tools they power, some readers worry they will be used to make creators redundant or automate book acquisitions in a sector already known for its lack of diversity.

# 56. [Artificial Intelligence (AI) Principles](https://www.wolterskluwer.com/en/about-us/artificial-intelligence-principles#:~:text=Wolters%20Kluwer%20focuses%20on%20privacy,through%20our%20processes%20and%20procedures.)

## AI Principles

Privacy and Security

* Wolters Kluwer focuses on privacy and security as part of the design, development and deployment of AI in our products and services. We promote the creation of AI systems that are safe, secure and reliable through our processes and procedures.

Transparency and Explainability

* Wolters Kluwer aims to design and develop AI systems with sufficient transparency and explainability to enable users to understand and use the system appropriately.

Governance and Accountability

* Wolters Kluwer adheres to development standards and processes that promote responsibility and accountability for AI systems and their outcomes. We address risk management and issue remediation during design and development, as well as after deployment.

Fairness

* Wolters Kluwer recognizes the importance of treating people fairly and without discrimination in the design and development of AI products and services.

Human Focused

* Wolters Kluwer strives to create AI systems that are human-centric, focused on solving business problems and benefiting our customers, while also considering the potential impact they may have on society and our environment.

# 57. [AI policy](https://www.onepercentfortheplanet.org/ai-policy)

# Artificial Intelligence Policy

1% for the Planet Artificial Intelligence Policy

Last updated: July 2023

Purpose

At 1% for the Planet, we recognize the potential benefits of artificial intelligence (AI) for supporting our stakeholders, our staff, and advancing our mission. We are committed to using AI in a responsible, ethical, and effective manner, and to fostering a culture of transparency and accountability in our use of AI tools.

To achieve these goals, we have developed this AI policy to guide our use of AI and ensure that it aligns with our [core values and mission](https://www.onepercentfortheplanet.org/about/mission). We will regularly review and update this policy to reflect changes in technology, industry standards, and legal and regulatory requirements. We subscribe to the “first draft not final draft” view of content or analyses generated by AI, and will always center our staff and stakeholders when considering the use of AI systems.

We are committed to being open to feedback and making changes as necessary to ensure that our use of AI is responsible, ethical, and mission-aligned. As the world of AI continues to evolve, we will continuously evaluate and improve upon our use of AI to ensure that it meets our high standards for ethical and responsible use.

Ethics & transparency

We will ensure that any AI systems or tools we deploy are guided by ethical principles, including an unwavering commitment to transparency in how we use AI. We will regularly review and update this publicly-available AI policy to reflect the latest ethical standards and best practices.

We will be transparent with our stakeholders about our use of AI. This includes regularly communicating with our stakeholders about the development and use of AI systems, as well as how those systems or tools make decisions and which data they use.

Any content or images generated by AI will always be reviewed and edited by human experts to ensure they are accurate, meet our ethical standards, and do not contain any objectionable or harmful content.

Images generated by AI will be accurately credited. We will respect the work of authors and artists when designing prompts to avoid directly using living artists in prompts.

Content developed with the support AI will utilize the same attribution process we would typically use for both internal and external facing content. For example, we don’t typically credit an author for a blog post, as many staff may have contributed to a given piece. When in doubt though, we will over communicate about our use of AI as an assistant in generating content.

Data privacy & security

We will use AI in ways that protect the privacy of individuals and businesses, and the security of their data.

We will ensure that any data we collect or use is handled in accordance with our [existing privacy policy](https://www.onepercentfortheplanet.org/privacy-policy), best practices and applicable regulations. Sensitive or personal data will always be anonymized, encrypted, or otherwise protected if used with AI tools.

Preventing discrimination & bias

We will use AI to promote diversity, equity, and inclusion in our global network. We will work to ensure that our AI tools and services are accessible and inclusive for all.

We will prevent and address bias and discrimination in our AI systems by using diverse datasets to build any AI systems, by leaning into diverse input and expert review of AI systems, and by regularly auditing and testing for bias and discrimination.

We recognize that AI tools, and in particular Large Language Model (LLMs) are only as good as the content they are trained on, which means they are subject to the inaccuracies and biases of the training content. These LLMs can also  “hallucinate” (i.e. generate factually inaccurate content that may sound correct). As such, we are fully committed to training our staff on appropriate use of LLMs and AI systems, and ensuring human oversight of any published content.

Human oversight & risk mitigation

We will consult with external experts as we develop and deploy any AI systems.

Before deploying any AI system or tool, we will conduct a risk assessment and develop a plan for mitigating and managing those risks.

We will ensure that there is human oversight of any AI systems by training staff on AI risk management, and teaching them how to monitor and regularly review those systems, and intervene when necessary.

As described above, we are committed to expert human review of any content generated with the help of AI systems, prior to sharing or publishing outside of our organization.

Regulatory compliance

We will consult with legal experts to ensure that any AI systems we develop or deploy are in compliance with applicable laws and regulations.

Training & education

We are committed to investing in AI prompt training and education for staff, including training on how to use and interpret AI systems, as well as training on the ethical considerations associated with the use of AI.

Basic AI guidelines will be required in alignment with this policy to ensure staff is trained appropriately. We also will not allow the use of prompts that mimic living creators like artists or authors.

Sustainability

We recognize the importance of minimizing the environmental impact of any AI systems we use or develop. We will account for increased energy consumption as part of our GHG Scope 3 measurement and offsetting practice

Review & updates

This policy will be reviewed annually, or as needed based on changes in technology, regulation, or organizational needs. Updates will be communicated to all users and appropriate training will be provided.

# 58. [Principles for AI](https://wan-ifra.org/2023/09/global-principles-for-artificial-intelligence-ai/)

A first of their kind, these pioneering [Global Principles](https://cdn.wan-ifra.org/wp-content/uploads/2023/09/06095924/20230905-Global-AI-Principles-Formatted.pdf) provide guidance for the development, deployment, and regulation of AI systems and applications to ensure business opportunities and innovation can thrive within an ethical and accountable framework.  The Global Principles for AI aim to ensure publishers’ continued ability to create and disseminate quality content while facilitating innovation and the responsible development of trustworthy AI systems.

Addressing critical dimensions relating to intellectual property, transparency, accountability, quality and integrity, fairness, safety, design, and sustainable development, the Global Principles on AI mark an unprecedented collaboration that safeguards the interests of content creators, publishers, and consumers alike.

In the Principles, the organizations call for the responsible development and deployment of AI systems and applications, stating that these new tools must only be developed under established principles and laws that protect publishers’ intellectual property, brands, consumer relationships, and investments. The Principles state explicitly that AI systems’ “indiscriminate misappropriation of our intellectual property is unethical, harmful, and an infringement of our protected rights.”

News/Media Alliance President and CEO Danielle Coffey stated, “These Global AI Principles demonstrate the widespread agreement of publishers worldwide that their intellectual property, which is the product of significant investments they have made in providing quality journalistic and creative content, should be recognized and respected. AI systems are only as good as the content they use to train them, and therefore developers of generative AI technology must recognize and compensate publishers accordingly for the tremendous value their content contributes to the development of these systems.”

Digital Content Next CEO Jason Kint stated, “For decades, our member companies have pursued opportunities to bring trusted news and entertainment to new platforms and new distribution channels enabled by the internet. We know from experience that principles like these are necessary to make certain those opportunities continue to proliferate and serve as a guidepost for businesses and policymakers who are wrestling with the ethical and legal questions surrounding AI.”

Among other things, the Global AI Principles stipulate that developers, operators, and deployers of AI systems should:

* Respect intellectual property rights protecting the organiszations’ investments in original content.
* Leverage efficient licensing models that can facilitate innovation through training of trustworthy and high-quality AI systems.
* Provide granular transparency to allow publishers to enforce their rights where their content is included in training datasets.
* Clearly attribute content to the original publishers of the content.
* Recognise publishers’ invaluable role in generating high-quality content for training, and also for surfacing and synthesizing.
* Comply with competition laws and principles and ensure that AI models are not used for anti-competitive purposes.
* Promote trusted and reliable sources of information and ensure that AI-generated content is accurate, correct and complete.
* Not misrepresent original works.
* Respect the privacy of users that interact with them and fully disclose the use of their personal data in AI system design, training, and use.
* Align with human values and operate under global laws.

### **Global AI Principles**

#### **Introduction**

AI developers and regulators have a unique opportunity to establish an ethical AI framework to boost innovation and create new business opportunities, while ensuring that AI develops in a way that is responsible and sustainable. To achieve this, it is essential that AI systems are trained on content and data which is accessed lawfully, including by appropriate prior authorisations obtained for the use of copyright protected works and other subject matter, and that the content and sources used to train the systems are clearly identified. This document sets out principles that the undersigned publisher organisations believe should govern the development, deployment, and regulation of Artificial Intelligence systems and applications. These principles cover issues related to intellectual property, transparency, accountability, quality and integrity, fairness, safety, design, and sustainable development.

The proliferation of **AI Systems, especially Generative Artificial Intelligence (GAI)**, present a sea change in how we interact with and deploy technology and creative content. While AI technologies will provide substantial benefits to the public, content creators, businesses, and society at large, they also pose risks for the sustainability of the creative industries, the public’s trust in knowledge, journalism, and science, and the health of our democracies.

We, the undersigned organisations, fully embrace the opportunities AI will bring to our sector and call for the responsible development and deployment of AI systems and applications. We strongly believe that these new tools will facilitate innovative breakthroughs when developed in accordance with established principles and laws that protect publishers’ intellectual property (IP), valuable brands, trusted consumer relationships, and investments. The indiscriminate appropriation of our intellectual property by AI systems is unethical, harmful, and an infringement of our protected rights.

Our organisations represent thousands of creative professionals around the world, including news, magazine, and book publishers and the academic publishing industry such as learned societies and university presses. Our members invest considerable time and resources creating high-quality content that keeps our communities informed, entertained, and engaged. These principles – applying to the use of our content to train and deploy AI systems, as they are understood and used today – are aimed at ensuring our continued ability to innovate, create and disseminate such content, while facilitating the responsible development of trustworthy AI systems.

#### Intellectual Property

**1) Developers, operators, and deployers of AI systems must respect intellectual property rights**, which protect the rights holders’ investments in original content. These rights include all applicable copyright, ancillary rights, and other legal protections, as well as contractual restrictions or limitations imposed by rightsholders on the access to and use of their content. Therefore, developers, operators, and deployers of AI systems—as well as legislators, regulators, and other parties involved in drafting laws and policies regulating AI—must respect the value of creators’ and owners’ proprietary content in order to protect the livelihoods of creators and rightsholders.

**2)** **Publishers are entitled to negotiate for and receive adequate remuneration for use of their IP.** AI system developers, operators, and deployers should not be crawling, ingesting, or using our proprietary creative content without express authorisation. Use of intellectual property by AI systems for training, surfacing, or synthesising is usually expressly prohibited in online terms and conditions of the rightsholders, and not covered by pre-existing licensing agreements. Where developers have been permitted to crawl content for one purpose (for example, indexing for search), they must seek express authorisation for use of the IP for other purposes, such as inclusion within LLMs. These agreements should also account for harms that AI systems may cause, or have already caused, to creators, owners, and the public.

**3) Copyright and ancillary rights protect content creators and owners from the unlicensed use of their content.** Like all other uses of protected works, use of protected works in AI systems is subject to compliance with the relevant laws concerning copyrights, ancillary rights, and permissions within protocols. To ensure that access to content for use in AI systems is lawful, including through appropriate licenses and permissions obtained from relevant rightsholders, it is essential that rightsholders are able effectively to enforce their rights, and where applicable, require attribution and remuneration.

**4) Existing markets for licensing creators’ and rightsholders’ content should be recognised.** Valuing publishers’ legitimate IP interests need not impede AI innovation because frameworks already exist to permit use in return for payment, including through licensing. We encourage efficient licensing models that can facilitate training of trustworthy and high-quality AI systems

#### **Transparency**

**5) AI systems should provide granular transparency to creators, rightsholders, and users.**It is essential that strong regulations are put in place to require developers of AI systems to keep detailed records of publisher works and associated metadata, alongside the legal basis on which they were accessed, and to make this information available to the extent necessary for publishers to enforce their rights where their content is included in training datasets. The obligation to keep accurate records should go back to the start of the AI development to provide a full chain of use regardless of the jurisdiction in which the training or testing may have taken place. Failure to keep detailed records should give rise to a presumption of use of the data in question. When datasets or applications developed by non-profit, research, or educational third parties are used to power commercial AI systems, this must be clearly disclosed so that publishers can enforce their rights. Where developers use AI tools as a component into the process of generating knowledge from knowledge, there should be transparency on the application of these tools, including appropriate and clear accountability and provenance mechanisms, as well as clear attribution where appropriate in accordance with the terms and conditions of the publishers of the original content. Without limiting and subject to paragraphs 6 and 9, AI developers should work with publishers to develop mutually acceptable attribution and navigation standards and formats. Users should also be provided with comprehensible information about how such systems operate to make judgments about system and output quality and trustworthiness.

#### **Accountability**

**6) Providers and deployers of AI systems should cooperate to ensure accountability for system outputs.**AI systems pose risks for competition and public trust in the quality and accuracy of informational and scientific content. This can be compounded by AI systems generating content that improperly attributes false information to publishers. Deployers of AI systems providing informational or scientific content should provide all essential and relevant information to ensure accountability and should not be shielded from liability for their outputs, including through limited liability regimes and safe harbours.

#### **Quality and Integrity**

**7) Ensuring quality and integrity is fundamental to establishing trust in the application of AI tools and services.**These values should be at the heart of the AI lifecycle, from the design and building of algorithms, to inputs used to train AI tools and services, to those used in the  practical application of AI. A fundamental principle of computing is that a process can only be as good or unbiased as the input used to teach the system (rubbish-in-rubbish-out). AI developers and deployers should recognise that publishers are an invaluable part of their supply chain, generating high-quality content for training, and also for surfacing and synthesising. Use of high-quality content upstream will contribute to high-quality outputs for downstream users.

#### **Fairness**

**8) AI systems should not create, or risk creating, unfair market or competition outcomes.**AI systems should be designed, trained, deployed, and used in a way that is compliant with the law, including competition laws and principles. Developers and deployers should also be required to ensure that AI models are not used for anti-competitive purposes. The deployment of AI systems by very large online platforms must not be used to entrench their market power, facilitate abuses of dominance, or exclude rivals from the marketplace. Platforms must adhere to the concept of non-discrimination when it comes to publishers exercising their right to choose how their content is used.

#### **Safety**

**9) AI systems should be trustworthy.** AI systems and models should be designed to promote trusted and reliable sources of information produced according to the same professional standards that apply to publishers and media companies. AI developers and deployers must use best efforts to ensure that AI generated content is accurate, correct and complete. Importantly, AI systems must ensure that original works are not misrepresented. This is necessary to preserve the value and integrity of original works, and to maintain public trust.

**10) AI systems should be safe and address privacy risks.** AI systems and models in particular should be designed to respect the privacy of users who interact with them. Collection and use of personal data in AI system design, training, and use should be lawful with full disclosure to users in an easily understandable manner. Systems should not reinforce biases or facilitate discrimination.

#### **By Design**

**11) These principles should be incorporated by design into all AI systems, including general purpose AI systems, foundation models, and GAI systems.**They should be significant elements of the design, and not considered as an afterthought or a minor concern to be addressed when convenient or when a third party brings a claim.

#### **Sustainable Development**

12) The multi-disciplinary nature of AI systems ideally positions them to address areas of global concern. AI systems bear the promise to benefit all humans, including future generations, but only to the extent they are aligned to human values and operate in accordance with global laws. Long-term funding and other incentives for suppliers of high-quality input data can help to align systems with societal aims and extract the most important, up-to-date, and actionable knowledge.

# 59. [AI and International security](https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/resources/docs/CNAS_AI%20and%20International%20Security.pdf)

NATIONAL SECURITY-RELATED APPLICATIONS OF ARTIFICIAL INTELLIGENCE Introduction There are a number of direct applications of AI relevant for national security purposes, both in the United States and elsewhere. Kevin Kelly notes that in the private sector “the business plans of the next 10,000 startups are easy to forecast: Take X and add AI.”1 There is similarly a broad range of applications for AI in national security. Included below are some examples in cybersecurity, information security, economic and financial tools of statecraft, defense, intelligence, homeland security, diplomacy, and development. This is not intended as a comprehensive list of all possible uses of AI in these fields. Rather, these are merely intended as illustrative examples to help those in the national security community begin to think through some uses of this evolving technology. (The next section covers how broader AI-driven economic and societal changes could affect international security.) Cybersecurity The cyber domain represents a prominent potential usage arena for AI, something senior leaders have expressed in recent years. In October 2016, National Security Agency (NSA) Director Michael Rogers stated that the agency sees AI as “foundational to the future of cybersecurity.” Rogers’ remarks occurred only two months after DARPA held its first Cyber Grand Challenge, a head-to-head fight between autonomous machines in cyberspace. Each system was capable of automatically discovering and exploiting cyber vulnerabilities in its opponents while patching its own vulnerabilities and defending itself from external cyberattacks.2 Impressed with the tournament’s results, DoD began a new program, Project Voltron, to develop and deploy autonomous cybersecurity systems to scan and patch vulnerabilities throughout the U.S. military.3 Even as DoD has begun to implement this technology, potential applications of AI for cybersecurity continue to evolve. The systems in the first Cyber Grand Challenge used rulebased programming and did not make significant use of machine learning. Were a similar competition to be held today, machine learning would likely play a much larger role. Below are several illustrative applications of machine learning in the cybersecurity domain that could be especially impactful for the international security environment. Increased Automation and Reduced Labor Requirements Cyber surveillance has tended to be less labor-intensive than the traditional human surveillance methods that it has augmented or replaced. The increased use of machine learning could accelerate this trend, potentially putting sophisticated cyber capabilities that would normally require large corporation or nation-state level resources within the reach of smaller organizations or even individuals.4 Already there are countless examples of relatively unsophisticated programmers, so-called “script kiddies,” who are not skilled enough to develop their own cyber-attack programs but can effectively mix, match, and execute code developed by others. Narrow AI will increase the capabilities available to such 4 4 actors, lowering the bar for attacks by individuals and non-state groups and increasing the scale of potential attacks for all actors. Using AI to Discover New Cyber Vulnerabilities and Attack Vectors Researchers at Microsoft5 and Pacific Northwest National Laboratory6 have already demonstrated a technique for using neural networks and generative adversarial networks to automatically produce malicious inputs and determine which inputs are most likely to lead to the discovery of security vulnerabilities. Traditionally, such inputs are tested simply by randomly modifying (aka “fuzzing”) non-malicious inputs, which makes determining those that are most likely to result in new vulnerability discovery inefficient and laborintensive. The machine learning approach allows the system to learn from prior experience in order to predict which locations in files are most likely to be susceptible to different types of fuzzing mutations, and hence malicious inputs. This approach will be useful in both cyber defense (detecting and protecting) and cyber offense (detecting and exploiting). Automated Red-teaming and Software Verification and Validation While there is understandable attention given to new vulnerability discovery, many cyber attacks exploit older, well-known vulnerabilities that system designers have simply failed to secure. SQL-injection, for example, is a decades-old attack technique to which many new software systems still fall prey. AI technology could be used to develop new verification and validation systems that can automatically test software for known cyber vulnerabilities before the new software is operationally deployed. DARPA has several promising research projects seeking to utilize AI for this function. Automated Customized Social Engineering Attacks Many major cybersecurity failures began with “social engineering,” wherein the attacker manipulates a user into compromising their own security. Email phishing to trick users into revealing their passwords is a well-known example. The most effective phishing attacks are human-customized to target the specific victim (aka spear-phishing attacks) – for instance, by impersonating their coworkers, family members, or specific online services that they use. AI technology offers the potential to automate this target customization, matching targeting data to the phishing message and thereby increasing the effectiveness of social engineering attacks.7 Moreover, AI systems with the ability to create realistic, low-cost audio and video forgeries (discussed more below) will expand the phishing attack space from email to other communication domains, such as phone calls and video conferencing.8 Information Security The role of AI in the shifting threat landscape has serious implications for information security, reflecting the broader impact of AI, through bots and related systems in the information age. AI’s use can both exacerbate and mitigate the effects of disinformation within an evolving information ecosystem. Similar to the role of AI in cyber attacks, AI provides mechanisms to narrowly tailor propaganda to a targeted audience, as well as 5 5 increase its dissemination at scale – heightening its efficacy and reach. Alternatively, natural language understanding and other forms of machine learning can train computer models to detect and filter propaganda content and its amplifiers. Yet too often the ability to create and spread disinformation outpaces AI-driven tools that detect it. Targeted Propaganda and Deep Fakes Computational propaganda inordinately affects the current information ecosystem and its distinct vulnerabilities. This ecosystem is characterized by social media’s low barriers to entry, which allow anonymous actors – sometimes automated – to spread false, misleading or hyper-partisan content with little accountability. Bots that amplify this content at scale, tailored messaging or ads that enforce existing biases, and algorithms that promote incendiary content to encourage clicks point to implicit vulnerabilities in this landscape.9 MIT researchers’ 2018 finding that “falsehood [diffuses] significantly farther, faster, deeper and more broadly” than truth on Twitter, especially regarding political news, further illustrates the risks of a crowded information environment.10 AI is playing an increasingly relevant role in the information ecosystem by enabling propaganda to be more efficient, scalable, and widespread.11 A sample of AI-driven techniques and principles to target and distribute propaganda and disinformation includes: • Exploitation of behavioral data – The application of AI to target specific audiences builds on behavioral data collection, with machine learning parsing through an increasing amount of data. Metadata generated by users of online platforms – often to paint a picture of consumer behavior for targeted advertising – can be exploited for propaganda purposes as well.12 For instance, Cambridge Analytica’s “psychographic” micro-targeting based off of Facebook data used online footprints and personality assessments to tailor messages and content to individual users.13 • Pattern recognition and prediction – AI systems’ ability to recognize patterns and calculate the probability of future events, when applied to human behavior analysis, can reinforce echo chambers and confirmation bias.14 Machine learning algorithms on social media platforms prioritize content that users are already expected to favor and produce messages targeted at those already susceptible to them.15 • Amplification and agenda setting – Studies indicate that bots made up over 50 percent of all online traffic in 2016.16 Entities that artificially promote content can manipulate the “agenda setting” principle, which dictates that the more often people see certain content, the more they think it is important.17 Amplification can increase the perception of significance in the public mind. Further, if political bots are “written to learn from and mimic real people,” according to computational propaganda researchers Samuel Woolley and Philip Howard, then they stand to influence the debate. For example, Woolley and Howard point toward the deployment of political bots that interact with users and attack political candidates, weigh in on activists’ behavior, inflate candidates’ follower numbers, or retweet specific candidates’ messaging, as if they were humans.18 Amplifying damaging or 6 6 distracting stories about a political candidate via “troll farms” can also change what information reaches the public. This can affect political discussions, especially when coupled with anonymity that reduces attribution (and therefore accountability) to imitate legitimate human discourse.19 • Natural language processing to target sentiment – Advances in natural language processing can leverage sentiment analysis to target specific ideological audiences.”20 Google’s offer of political interest ad targeting for both “left-leaning” and “right-leaning” users for the first time in 2016 is a step in this direction.21 By using a systemic method to identify, examine, and interpret emotional content within text, natural language processing can be wielded as a propaganda tool. Clarifying semantic interpretations of language for machines to act upon can aid in the construct of more emotionally relevant propaganda.22 Further, quantifying user reactions by gathering impressions can refine this propaganda by assessing and recalibrating methodologies for maximum impact. Private sector companies are already attempting to quantify this behavior tracking data in order to vector future microtargeting efforts for advertisers on their platforms. These efforts are inherently dual-use – instead of utilizing metadata to supply users with targeted ads, malicious actors can supply them with tailored propaganda instead. • Deep fakes – AI systems are capable of generating realistic-sounding synthetic voice recordings of any individual for whom there is a sufficiently large voice training dataset.23 The same is increasingly true for video.24 As of this writing, “deep fake” forged audio and video looks and sounds noticeably wrong even to untrained individuals. However, at the pace these technologies are making progress, they are likely less than five years away from being able to fool the untrained ear and eye. Countering Disinformation While no technical solution will fully counter the impact of disinformation on international security, AI can help mitigate its efficiency. AI tools to detect, analyze, and disrupt disinformation weed out nefarious content and block bots. Some AI-focused mitigation tools and examples include: • Automated Vetting and Fake News Detection – Companies are partnering with and creating discrete organizations with the specific goal of increasing the ability to filter out fake news and reinforce known facts using AI. In 2017, Google announced a new partnership with the International Fact-Checking Network at The Poynter Institute, and MIT’s the Fake News Challenge resulted in an algorithm with an 80 percent success rate.25 Entities like AdVerif.ai scan and detect “problematic” content by augmenting manual review with natural language processing and deep learning.26 Natural language understanding to train machines to find nefarious content using semantic text analysis could also improve these initiatives, especially in the private sector. 7 7 • Trollbot Detection and Blocking – Estimates indicate the bot population ranges between 9 percent and 15 percent on Twitter and is increasing in sophistication. Machine learning models like the Botometer API, a feature-based classification system for Twitter, offer an AI-driven approach to identify them for potential removal.27 Reducing the amount of bots would de-clutter the information ecosystem, as some political bots are created solely to amplify disinformation, propaganda, and “fake news.”28 Additionally, eliminating specific bots would reduce their malign uses, such as for distributed denial-of-service attacks, like those propagated by impersonator bots throughout 2016.29 • Verification of Authenticity – Digital distributed ledgers and machine speed sensor fusion to certify real-time information and authenticity of images and videos can also help weed out doctored data. Additionally, blockchain technologies are being utilized at non-profits like PUBLIQ, which encrypts each story and distributes it over a peer-to-peer network to attempt to increase information reliability.30 Content filtering often requires judgement calls due to varying perceptions of truth and the reliability of information. Thus, it is difficult to create a universal filter based on purely technical means, and it is essential to keep a human in the loop during AI-driven content identification. Technical tools can limit and slow disinformation, not eradicate it. Economic and Financial Tools of Statecraft Illicit funds course through the global financial system and support terrorism, money laundering, and WMD proliferation. To counter these flows, U.S. officials have expanded the global network of anti–money laundering and counterterrorist financing tools since 9/11. Yet the United Nations estimates that law enforcement only seizes 1 percent of criminal funds.31 One potential national security application of AI tools is their use to strengthen counter–illicit–financing operations. By analyzing and learning from large sets of data, AI could accomplish tasks not possible in a human-centered counter–illicit–financing system. AI’s anomaly detection and pattern recognition capabilities could help a system learn from the unstructured data collected by financial institutions. In one case, a regulatory technology company integrating AI tools found a correlation between users who had changed their browser language and a type of fraud.32 This analysis uncovered a metric not traditionally used by financial investigators and expanded the definition of usable data. Better pattern recognition will also sort information more usefully. Better sorting can reduce false positives that would otherwise result in alerts. For example, AI could reduce false positives in “high-risk” jurisdictions by replacing an imprecise geographic input with a more effective red flag. Fewer alerts will save time and manpower. Even short of large-scale pattern analysis, AI can improve the counter–illicit–financing framework. Automation could ensure sustained attention to illicit financing threats, even when not prioritized by financial institutions. This feature would allow constant pressure 8 8 on potential dangers. It would also reduce stress on financial institutions. Banks would no longer have to shift their attention to respond to changing government priorities – for example, from Iran to North Korea. Automation could also integrate available non-financial information about entities and individuals. Today, a significant amount of publicly accessible information is not automatically part of investigations. Through image recognition, AI programs could use open-source social media information that currently does not inform counter–illicit–financing processes.33 Changes made to a customer’s social media presence or networks mapped out through publicly available images could clarify a customer’s risk profile.34 AI capabilities could address the counter–illicit–financing framework’s major challenges. First, AI could improve efficiency. Human-centered counter–illicit–financing processes generate false positives that detract from investigations and allow threats to go undiscovered or uninvestigated. One study found that 80 to 90 percent of suspicious activity reporting yielded no value.35 Fewer false positives, through better pattern recognition and data segmentation, will save time and money. The savings of time and money that AI systems could enable would be particularly important in combating illicit financial flows. Since 9/11, governments have enlisted financial institutions as partners in the fight against illicit finance. Banks have shouldered increasing compliance costs to keep up with growing regulatory requirements, including to counter illicit finance. Between 2011 and 2017, the cost of compliance has increased by over 20 percent for most banks.36 Lower costs will ensure banks’ continued cooperation. Lower costs will also allow smaller, regional banks in high-risk jurisdictions to conduct compliance work that currently only large multinational institutions can afford. Greater participation from smaller banks will reduce vulnerable entry points into the global financial system. AI could also help governments and financial institutions address data privacy and protection problems. Currently, privacy laws hamper efforts to make the most of collected information. In some cases, financial institutions can struggle even to share information among their own branches in different jurisdictions.37 These limitations create barriers to integrating information and, more importantly, to learning from past typologies of illicit financing. Dr. Gary Shiffman, CEO of Giant Oak, a data science company that uses algorithms to understand large quantities of data, argues that AI could circumvent this problem. An AI system could learn from analyzing a dataset in one jurisdiction. The system could then move its algorithms to other jurisdictions and learn from a new dataset without moving the underlying data itself.38 Privacy limitations would no longer hamper the learning. Though AI could incrementally improve the counter–illicit–financing framework, it could also fundamentally disrupt it. Financial institutions often use static rules to counter illicit funding. For example, a transaction over $10,000 will trigger a currency transaction report. Rogue players, however, can adapt faster than the rules can evolve. For this reason, international standard-setters like the Financial Action Task Force (FATF) urge financial institutions to use risk-based systems that proactively adapt to and mitigate risk. Because 9 9 this approach is costly and time-intensive, FATF requires risk-based measures to tackle money laundering and terrorist financing, but not for the financing of proliferation. An AIbased system, constantly learning and incorporating new information, will allow the expansion of risk-based programs to create a more dynamic counter–illicit–finance program across threat categories.39 An AI system, for example, could spot the patterns used by individuals to evade the $10,000 limit, connect these illicit networks, and potentially block the wires from leaving banks before transferred amounts become too big.40 This section has focused on applying AI to flows of finance rather than the infrastructure and markets supporting these flows. The “flash crash” has shown the susceptibility of programmed trading mechanisms to negative interactions and the currently insufficient preparation for this threat.41 Opponents could use AI to manipulate markets or destabilize currencies. This category of threats, however, falls outside the traditional realm of economic statecraft. Instead, it would be analogous to a malicious cyber attack. Making the most of financial data will be particularly important going forward. As more and more communication becomes encrypted, financial records will become more important sources of data for investigations and intelligence work. However, the tools to use the data have not yet evolved accordingly. AI offers a way forward. Defense Militaries around the globe are already incorporating more robotics and autonomous systems into their forces, a trend that has been the focus of prior CNAS work. Artificial intelligence and machine learning will allow these systems to tackle more challenging tasks in a wider range of environments. Because of the ubiquitous nature of AI technology, nonstate groups and individuals will also be able to harness and use this technology. In combat operations, robots, swarms, and autonomous systems have the potential to increase the pace of combat. This is particularly the case for domains of machine-tomachine interaction, such as in cyberspace or the electromagnetic spectrum. AI could be used not only to create more intelligent robotics, but also to power more advanced sensors, communications, and other key enablers. • Situational awareness: Small robotic sensors could be used to collect information, and AI-enabled sensors and processing could help make better sense of that information. Deep neural networks already are being used for image classification for drone video feeds as part of the Defense Department’s Project Maven, in order to help humans process the large volumes of data being collected. While current AI methods lack the ability to translate this into an understanding of the broader context, AI systems could be used to fuse data from multiple intelligence sources and cue humans to items of interest. AI systems also could be used to generate tailored spoofing attacks to counter such sensors and processors. 10 10 • Electromagnetic spectrum dominance: AI systems could be used to generate novel methods of jamming and communications through self-play, akin to AlphaGo Zero improving its game by playing itself. For example, one AI system could try to send signals through a contested electromagnetic environment while another system attempts to jam the signal. Through these adversarial approaches, both systems could learn and improve. DARPA held a Spectrum Challenge in 2014 with human players competing to send radio signals in a contested environment.42 DARPA is now using machine learning to aid in radio spectrum allocation,43 but this concept also could be applied to jamming and creating jam-resistant signals. • Decoys and camouflage: Generative adversarial networks could be used to create militarily relevant deep fakes for camouflage and decoys, and small robotic systems could be used as expendable decoys. As militaries incorporate more AI-enabled sensors for data classification, spoofing attacks against such systems will be increasingly relevant as well. • Tactics: Evolutionary and reinforcement learning methods could be used to generate new tactics in simulated environments, coming up with surprising solutions as they have in other settings. • Command and control: As the pace of battle accelerates and the volume and speed of information eclipses the ability of human warfighters, AI will become increasingly important for command and control. Autonomous systems that have been delegated authority for certain actions can react at machine speed at the battlefield’s edge without waiting for human approval. AI can also help commanders process information faster, allowing them to better understand a rapidly changing battlespace. Through automation, commanders can then relay their orders to their forces – human or machine – faster and more precisely. AI systems can also aid militaries in a range of non-combat support functions. One use of AI will be to help defense leaders better understand their own forces. By analyzing large amounts of data, AI systems may be able to predict stress on the force in various components: when equipment requires maintenance; when programs are likely to face cost overruns or schedule delays; and when servicemembers are likely to suffer degraded performance or physical or psychological injuries. Overall, AI has tremendous potential to help defense leaders improve the readiness of their own forces by assembling and fusing data and doing predictive analysis so that problems can be addressed before they become critical. AI also is ripe for transforming traditional business processes within military and other government organizations. The U.S. Defense Department, for example, conducts a range of non-military specific business functions, including accounting, travel, medicine, logistics, and other administrative functions. Many of these functions are ripe for automation because they involve routine cognitive or physical labor. In many cases, military organizations may be able to directly import mature and proven technologies from the 11 11 commercial sector that can improve efficiencies and reduce personnel costs, such as more automated accounting systems or AI tools in health care. Defense organizations could save substantial sums of money by drawing on these commercial technologies and streamlining their organizations. Overall, artificial intelligence can help militaries improve understanding, predict behavior, develop novel solutions to problems, and execute tasks. Some applications, such as the use of AI to enable autonomous weapons, raise difficult legal, ethical, operational, and strategic questions. The potential for automation to increase the pace of combat operations to the point where humans have less control over the conduct of war raises profound questions about humanity’s relationship with war, and even the nature of war itself. Intelligence AI has many uses in intelligence collection and analysis. For collection, the explosion of data that is occurring because of smart devices, the Internet of Things, and human internet activity is a tremendous source of potential information. This information would be impossible for humans to manually process and understand, but AI tools can help analyze connections between data, flag suspicious activity, spot trends, fuse disparate elements of data, map networks, and predict future behavior. This could make clandestine activity more challenging in a number of ways, as the combination of big data, data breaches, and increased open source information could make it more difficult to keep intelligence professionals undercover. For example, facial recognition and biometrics, combined with large surveillance systems, could make operating under aliases increasingly difficult. At the same time, AI systems may be vulnerable to counter-AI spoofing techniques, such as fooling images, which will have implications for the intelligence community. Deep fakes and the automation of data creation at scale may make it possible to create deep backstories for individuals undercover. AI may even transform verification of human reporting through improvements in systems that can correlate brain imaging to thoughts, with major implications for counter-intelligence and interrogation.44 AI also has tremendous potential value in intelligence analysis. AI systems can be used to track and analyze large amounts of data – including open-source data – at scale, looking for indications and warning of suspicious activity. Anomaly detection can help find terrorists, clandestine agents, or indications and warning of potential enemy military activity. AIbased speech-to-text and translation services could greatly increase the scale of processing audio, video, and text-based foreign language information. AI systems could be used to generate simple automated reports, as they do already for some sports games.45 AI systems generally perform poorly at reading comprehension, but as they improve they could be used increasingly to write summaries of transcripts, making it easier for human analysts to quickly sift through the ever-growing volumes of information.46 AI systems also could be increasingly valuable in doing semantic analyses of reports that help link disparate pieces of data that humans might miss. AI systems lack the common-sense reasoning that would allow them to make sense of information, but their ability to operate 12 12 with precision at scale will aid human analysts in sorting through massive volumes of information. AI systems will not replace human intelligence analysts, but can aid them by offloading routine tasks and processing data at scale, allowing human analysts to focus on understanding adversaries. Homeland Security AI can also aid a variety of border security and homeland security applications. AI-driven perception, processing, and analysis will be essential for collecting, sorting, and interpreting data to better inform human decision-making. The U.S. Department of Homeland Security (DHS) has already started to adopt and implement some of these technological advancements. Examples of past and current AI-driven DHS initiatives include: • Voice recognition algorithms – The U.S. Coast Guard has used artificial intelligence to analyze voices to build out their physical appearances. This has helped forensically address false distress signals.47 • Open source data for machine learning – In conjunction with Alphabet Inc.’s Kaggle platform, DHS made data from the Transportation Security Administration available to develop better algorithms to evaluate passenger luggage for illicit and dangerous items.48 • Understanding data – The Assistant for Understanding Data through Reasoning, Extraction, and Synthesis (AUDREY) AI platform developed by DHS and NASA’s Jet Propulsion Laboratory integrates real-time data to make recommendations to firefighters on how to best function as a team.49 AI also has broad applicability in a variety of homeland security functions, such as border security.50 Since the U.S. government cannot be stationed at every mile or inspect every container, AI systems, potentially in combination with UAVs and ground robotics, can aid in monitoring borders through advances in automated surveillance and anomaly detection. Systems that monitor human emotional expression and behavior could aid in recognizing humans that appear nervous or are acting oddly, serving as a “sixth sense” at border crossings. AI systems used for game theory/risk assessment also could be valuable in determining where best to apply scarce resources and how to counter adaptive adversaries, such as drug traffickers. Indeed, such systems already are being used to improve security against poachers in Africa.51 Diplomacy and Humanitarian Missions Advances in artificial intelligence could also reshape the practice of diplomacy. AI technologies in image recognition and information sorting can make diplomatic compounds safer by monitoring personnel and identifying anomalies for potential vulnerabilities. In addition, language processing algorithms will lower language barriers 13 13 between countries, allowing them to communicate to foreign governments and publics more easily. More theoretical technologies like political forecasting also remain an option, mining an increasing array of available data to better understand and predict political, economic, and social trends.52 However, diplomacy will not be without disruptive challenges. Humans, for the foreseeable future, remain the decision-makers and must properly use the outputs provided by AI technologies. More alarmingly, as efforts to forge testimonies and propagate disinformation abroad are made easier, AI technologies will have to be applied defensively to react to, correct, or even remove malicious content.53 International humanitarian operations could also benefit greatly from AI technologies. AI technologies can help monitor elections, assist in peacekeeping operations, and ensure financial aid disbursements are not misused through anomaly detection. Of course, artificial intelligence can also help directly improve the quality of life in less developed nations by increasing productivity, health care, and myriad other economic benefits.54 Artificial intelligence could also help in avoiding disasters that lead to international intervention. For example, AI technologies that extract significant actionable warning signs from climate and soil patterns will be a boon in agricultural efficiency and disaster preparedness.55 Implications As an enabling technology, AI has many uses across a variety of national security settings. The United States should expand upon nascent efforts within different parts of the government and establish a whole-of-government initiative to harness and rapidly integrate AI tools within government operations. Because many current AI approaches have significant vulnerabilities, the United States should include safety and robustness against adversarial manipulation as key elements of its effort to incorporate AI technology, and employ “red teams” to test AI tools before they are deployed. The ubiquitous nature of AI technology means that the United States will have to move quickly to keep ahead of potential competitors. 14 14 THE INDIRECT EFFECTS OF THE ARTIFICIAL INTELLIGENCE REVOLUTION FOR GLOBAL SECURITY How might AI generate political and societal change relevant for the international security environment beyond direct national security implications? Given the integral link between economic and military power, especially over the medium to long run, understanding how AI innovations will shape the global economy, the information environment, and societies around the world is crucial. Economic Power and the Future of Work The clearest connections between AI, the global economy, and economic power are through the effect of AI on the ability of countries and businesses to accumulate capital and the consequences for the future of work. The question is whether the consequences of AI will match, or even exceed, previous large-scale shifts in the economy. For example, in 1820 71 percent of Americans reportedly worked in farming occupations. However, the percentage of Americans working in farming declined significantly over the next century due to industrialization, falling to 30 percent in 1920 and 1 percent by 1988.56 There has been a large range of predictions on the way that AI will shape the labor market, and those predictions have a large degree of uncertainty. For example, a recent report by the McKinsey Global Institute suggests that nearly half of current job tasks across industries are automatable, while in six out of ten jobs more than 30 percent of the job tasks are automatable. The midpoint estimate of number of jobs displaced by 2030, according to McKinsey, is 400 million, while the high-end estimate is twice as high – 800 million.57 These enormous totals, and the wide spread between them, reflect not just the notion that AI will have significant consequences on the labor market, but that those consequences are difficult to predict. Researchers’ estimates on the effect of automation vary significantly. Research by Carl Benedict Frey and Michael A. Osborne at Oxford University suggests that 47 percent of U.S. workers might be at risk from automation by about 2030. 58 Another report examining 32 developed countries in the Organisation for Economic Cooperation and Development argues that 14 percent of jobs are at a high risk of automation and another 32 percent of jobs are at significant risk.59 Meanwhile, a U.S. labor market model by Daron Acemoglu and Pascual Restrepo at the National Bureau of Economic Research, based on data on industrial robotics from 1990–2007, suggests adding “one more robot in a commuting zone reduces employment by 6.2 workers.”60 A Forrester research report, in contrast, argues that only 24.7 million jobs will be displaced by 2027, with 14 million created.61 And even McKinsey says that only about 5 percent of jobs as they exist today could be fully automated.62 However, this is not just a question of how many jobs are displaced versus created, but whether those displaced will be able to find work in the new economy. The process of creative destruction can have significant political consequences even if the macro economic effects are relatively stable.63 Former Secretary of the Treasury Larry Summers argued in 2017 that automation pressures, in combination with the difficulty of generating new skills 15 15 for labor force participation later in life, could result in “a third of men between the ages of 25 and 54 not working by the end of this half century.”64 The challenge is that the number of jobs created by the cutting-edge companies of today, at the outset of the AI revolution, is already much smaller than the number of jobs created by the leading companies of previous generations. For example, in 2017 Facebook employed a little over 25,000 people, the largest it has ever been. Meanwhile, Ford Motor Company, with a fraction of the size of its peak labor force, still employed 202,000 workers in 2017. 65 The risk is that the optimal economic future for growth is more of a “labor-light economy,” as Erik Brynjolfsson and Andrew McAfee argue, where capital generates continuing productivity gains, but workers don’t benefit. 66 And workers, in this scenario, would not be just factory workers. They would be lawyers, doctors, investment bankers, and others that currently have middle class, upper middle class, or upper class incomes. All of those jobs have repetitive tasks, no matter how skilled, that narrow autonomous systems may be able to master. In this scenario, workers who perform repetitive physical and cognitive labor become less valued. Even if unemployment is low, reduced wages can be the effect. In fact, Brynjolfsson and McAfee argue that automation has been responsible for stagnant or falling real wages for the median American worker for the past several decades.67 Eras with this level of disruption can have significant indirect implications for the balance of power and the security environment. A change in the underlying basis of the economy can lead to industry shifts that benefit some countries at the expense of others. For example, the First Industrial Revolution helped fuel the rise of the United States – the geography of the United States enabled industrialization on a scale that was difficult to achieve in Europe. Government policy to capitalize on these changes can lead to longlasting shifts in the relative balance of power. The ability of the British government to establish modern financing, in terms of government borrowing and bond markets, enabled Great Britain’s creation of the most powerful navy in the world in the late 19th century.68 Political and Social Disruption Economic disruption can also fuel social and political disruption. Large numbers of formerly employed workers, or even just groups that are newly disadvantaged due to economic circumstances, are a recipe for political protest and agitation. Maintaining stability requires a level of political dexterity and bureaucratic competence that can be difficult to achieve at the best of times – and periods of economic instability are hardly the best of times. This is one of the mechanisms through which economic transitions can lead to political conflict that, in the worst case, can make domestic unrest, insurgencies, civil wars, nationalism, xenophobia, and a turn to authoritarianism more likely. The instability generated by automation is already a potential driving force in the rise of populist nationalist movements around the world. As powerful interest groups such as coal workers experience significant decline, they become ever more radical in their desire to see change to return to an old status quo that is impossible to achieve. This can drive political polarization. A disparity in how automation affects different demographic groups could conceivably drive internal political conflict. To better understand how automation would affect American workers, the authors compared the analysis done by the McKinsey Global Institute on the effects of automation by sector69 to the age of workers in each sector, as identified by the Bureau of Labor Statistics.70 Figure 1 below shows the results.71 The results suggest that automation is likely to hit younger workers the hardest in the United States. This is not surprising, as younger workers are the most likely to be performing routine tasks that are easily automatable. This is particularly true for workers aged 16–19 and 20–24, who are less likely to be highly educated. Worker wages sharply decoupled by education level in the 1980s, with inflation-adjusted wages for those with a college or post-graduate degree rising and wages for high school graduates and dropouts falling.72 This suggest that one effect of the automation economy will to be magnify the impact of education even more – even of specific majors or disciplines that help prepare people for the jobs of the future. From the standpoint of managing the consequences of creative destruction, the silver lining is that the workers hardest hit by automation are those who are youngest and have the 17 17 most time to gain an education and adapt. One risk, however, is that younger workers who rely on entry-level jobs to pay their way through college and obtain an education could lose the economic opportunities they need to stay relevant in the automation economy. Without policy adjustments to make college and post-graduate education more affordable, the result therefore could be rising inequality. National Economic Scenarios Regarding AI Governments are not passive players at the mercy of a tidal wave of automation. Nations have a range of policy options available for responding to the economic pressure that AI will likely generate, from regulating industries to introducing a universal basic income. Countries will undoubtedly want to take advantage of AI as best they can while minimizing its harmful effects. What form that takes will depend on each country’s political economy and the national attitudes toward economic growth, unemployment, political unrest, and social welfare. The consequences of AI plus national policy responses could vary widely. Below are a few illustrative scenarios for how nations might end up after weathering and adapting to a wave of AI-driven creative disruption. • Bounty – The advantages of AI in increasing productivity and prosperity could vastly outweigh the disadvantages to workers, and the outcome could be wealth and abundance for all, even those displaced by automation. • Rising inequality – Even if workers displaced by AI find new jobs, the result could be rising inequality in a labor-light economy, as capital becomes more valuable and the wealthy get wealthier. As inequality widens, social and political instability could result. • Resource curse – AI could lead to an economic paradox, much like the “resource curse” faced by countries abundant in natural resources. Even policy measures like universal basic income could fail to effectively translate to societal well-being and individual happiness. • Luddite’s revenge – A dire scenario could be massive unemployment, as the fears of the 19th century Luddites finally come true and machinery eliminates jobs that are not replaced by new ones. One effect of narrow AI could be that humans simply are not as economically valuable as they once were, much like the decline in the role of horses in the global economy following the first and second industrial revolutions.73 • Generational dislocation – Like the move from the field to the factory, AI could cause a transformation in the labor market that takes a generation to resolve. With a fundamental skills mismatch between the people who have lost jobs and the skills needed for new jobs created by AI, the result could be social and political 18 18 disruption lasting a generation. This disruption resolves itself over time as a new generation, educated and trained in the AI economy, dominates the labor market. • Fall behind – Nations that fail to take advantage of AI or even resist it, for fear of potential economic and political disruption, could fall behind other countries, maintaining stability but at the cost of growth and national competitiveness. Universal Basic Income Fear of the large-scale dislocation potential of AI, and the enormous social and political consequences that result, is a large driver of recent discussions about the possibility of universal basic income. Universal basic income represents the idea that the government would provide income, sufficient to live on, for everyone. High-profile business leaders such as Richard Branson have argued that universal basic income might become a necessity due to AI.74 Essentially, if the labor market implications of AI are such that new industries and possibilities for human work do not emerge, huge segments of the population could end up more or less out of work, with capital concentrated ever more in the hands of the ultra-wealthy. This would not necessarily be due to corruption or poor decision-making, just the logic of the marketplace taken to an extreme. Thus, one potential solution is to offer those who are displaced by automation the potential for a guaranteed income, given that they are unlikely to have future workplace options.75 Universal basic income raises many questions, of course. Who is paying in for universal basic income, and on what basis? Moreover, what about the possibility for adverse incentives? Universal basic income would essentially lessen the cost of free-riding on the system. It is also possible that universal basic income could reduce the incentive for innovation among people who otherwise would work hard to find new, productive industries where humans would have a comparative advantage over machines in an era of artificial intelligence. These are hard questions, and ones that policymakers will have to consider over the next decades. Nationalism and International Conflict As described elsewhere in this report, the clearest national security consequence related to the economics of AI will be the integral link between economic power and military power. It is simply not possible to maintain a leading military over time with a declining economy. The analysis above also suggests, however, that the economic, social, and political dislocation caused by AI could generate additional international security consequences. Today, there are already political pressures in Western countries such as the United States and Great Britain that are focused on the ways the countries have changed for the worse. Automation and artificial intelligence have not yet received the blame for this, interestingly, despite the evidence presented above about the impact that automation has already had on 19 19 the labor market. Instead, political arguments in the West often focus on issues such as immigration, outsourcing, or trade deficits with countries such as China.76 If job losses, or even just labor force instability, from artificial intelligence accelerate, it could unleash a larger wave of populism and nationalism, as wealth concentration in the hands of a smaller and smaller number of elites generates resentment and political instability. On the global stage, labor force instability at the level AI could generate has in the past led to mass turmoil, coups, and other tension, as well as the type of virulent nationalism that can generate conflict, particularly if populations blame other nations for their economic woes. The Information Environment Digital technologies have radically transformed the information environment in the span of only a few decades, democratizing the number of voices, expanding the volume, and accelerating the speed of societal discourse. AI will continue to change the information environment as computers become more capable of targeting information at specific users, amplifying messages, filtering information, and even generating fake audio, images, and videos. The rapid evolution of the internet, social media, and disinformation suggests it is impossible to predict how the information environment will evolve. Below are some challenges, however, that one can anticipate based on existing technology. The End of Truth AI has already demonstrated the ability to create audio and visual forgeries. Dr. Hany Farid, a professor of computer science at Dartmouth University who consults for the Associated Press to detect forged images and other media, has described the competition between forgery technology and authentication technology as an “arms race” and an “information war.”77 At the moment, recording and authentication technology has the upper hand, but the trends are not favorable. Society may be only a few years away from such forgeries being able to fool not just the untrained eye and ear, but sophisticated forgery detection experts and systems.78 This shift will bring profound implications across domains as diverse as corporate communications, courtroom evidence, journalism, and international security. Take, for instance, the Watergate scandal. President Richard Nixon maintained sufficient support in the Senate to block his removal from office even after two years of aggressive investigative reporting. Only upon the release of the “smoking gun” Oval Office audiotapes – where Nixon can be heard explicitly condoning a criminal cover-up and obstruction of justice – did his support in Congress finally fail. In a world where realistic forgeries were essentially impossible, audiotapes served not just as evidence but as undeniable proof. AI technology could weaken, if not end, recorded evidence’s ability to serve as proof. Some technologies, such as blockchain, may make it possible to authenticate the provenance of video and audio files. These technologies may not mature quickly enough, though. They could also prove too unwieldy to be used in many settings, or simply may not be enough to counteract humans’ cognitive susceptibility toward “seeing is believing.” The result could 20 20 be the “end of truth,” where people revert to ever more tribalistic and factionalized news sources, each presenting or perceiving their own version of reality. AI-enabled forgeries are becoming possible at the same time that the world is grappling with renewed challenges of fake news and strategic propaganda. During the 2016 U.S. presidential election, for example, hundreds of millions of Americans were exposed to fake news. The Computational Propaganda Project at Oxford University found that during the election, “professional news content and junk news were shared in a one-to-one ratio, meaning that the amount of junk news shared on Twitter was the same as that of professional news.”79 A common set of facts and a shared understanding of reality are essential to productive democratic discourse. The simultaneous rise of AI forgery technologies, fake news, and resurgent strategic propaganda poses an immense challenge to democratic governance. Political Power, Democracy, and Authoritarianism Due to private and public actors’ ability to use AI techniques to shape information flows and perceptions, they could affect democratic processes and the strength of authoritarian regimes while also shaping global public discourse. Potential application areas include: • Electoral process influence – Highly granular voter profiling, enabled by the application of AI technologies, can affect democratic norms through the electoral process. Certain advances are likely to see more narrowly targeted content creation, with bots used to amplify this messaging in targeted sub-groups. For instance, these technologies were used in targeted political ads based on the social media profiles of voters in the 2016 U.S. presidential election and the U.K. Brexit referendum.80 Mitigation measures for this personalized propaganda – often in private messages so no public data can be gathered and scrutinized – include Facebook’s pledge to make all “dark ads” on its platform public.81 • Authoritarian regimes – Social media allows authorities to manipulate the news environment and control messaging. In China today, reports estimate that the government creates and posts about 448 million social media comments a year.82 In some cases, bots are utilized to run propaganda efforts both inside and outside a home country, with the aim of creating a strategic advantage in today’s crowded information ecosystem.83 • Social media – The nature of AI makes it liable to concentrate information influence in the hands of a limited number of media platforms. Private companies not only control the data they collect, but can actively promote and demote specific content. Google, for example, de-ranks specific news outlets in its search results and only includes “publishers that are algorithmically determined to be an authoritative source of information” in its fact-checking features.84 21 21 • Future targeting efforts – Uses of AI to target audiences and spread disinformation include the expansion of automated spear phishing and sophisticated targeting of public sector employees with intent to influence government operations (i.e., orders spoofing).85 Actors could also use AI to create “automated, hyper-personalized disinformation campaigns,” in which certain key demographics or areas (i.e., swing districts) are targeted to affect voting behavior at crucial times, potentially resulting in election shaping through sophisticated AI systems.86 Because AI tools can be deployed at scale without large numbers of people, these tools could enable small numbers of people to wield outsize political influence, whether through governments, corporations, or other groups. The effect could be to erode the power of the people and democratic institutions and enable new forms of authoritarianism. CONCLUSION What world do we end up in? Does AI usher in a new era of prosperity and international peace? Does it lead to shifts in the balance of power on the global stage, with attendant risks of conflict and miscalculation? Could AI lead to massive dislocation and a rise in political unrest, nationalism, and protectionism? Does AI concentrate power to control information in the hands of a few, or continue the democratization of information that computers, networks, and social media have unleashed? Does the cacophony of competing information lead to a turn away from truth to authoritarianism and tribalism, or does the wisdom of the crowds win out with a convergence on truth and centrist policies? The technological opportunities enabled by artificial intelligence shape the future, but do not determine it. Nations, groups, and individuals have choices about how they employ and respond to various uses of AI. Their policy responses can guide, restrict, or encourage certain uses of AI. In order to manage the challenges ahead, the United States will need to adopt a national strategy for how to take advantage of the benefits of AI while mitigating its disruptive effects.