ASG ANALYSIS:

The Geopolitics of Generative AI

Key takeaways

- Rapid advances in the capabilities of generative artificial intelligence (AI) systems that can produce natural-sounding conversations, novel images, and computer code are reverberating through the technology sector and geopolitics.

- Governments and companies are scrambling to understand the policy and regulatory implications of these systems. This is happening in a geopolitical environment characterized by intensifying U.S.-China tech competition; European concerns about over-reliance on U.S. technology platform companies; and efforts to develop regulatory strategies to ensure the safety, fairness, and reliability of AI.

- Chinese tech companies will have an advantage building generative AI models that are optimized for their domestic market, but will also face domestic political challenges as Beijing attempts to ensure that chatbots and other services enabled by the technology do not undermine Chinese Communist Party authority. U.S. export controls on cutting-edge semiconductors may also limit Chinese firms' ability to develop and commercialize these systems.

- In Europe, big U.S. cloud services companies' investments in AI startups and moves to incorporate chatbots and other generative AI capabilities into search engines will fuel concerns about digital sovereignty. Generative AI will also add new urgency to European debates about competition policy, AI regulation, and content moderation.

- In the U.S., China hawks may seize on generative AI to push for tighter curbs on U.S. outbound investment in China's technology sector, tougher export controls, and restrictions on China's access to data sets or application programming interfaces used by AI developers to deliver innovative new services. State legislatures will be a more likely source of new legislation targeting AI than the U.S. Congress.

- To reduce the chance of a regulatory backlash or other political disruption of the sector, companies developing and deploying generative AI systems and their supporting technology ecosystem should closely monitor these trends and seek ways to educate policymakers and the public about the capabilities and limits of the technology.

Excitement and concern over AI's growing capabilities

Recent advances in generative AI – an industry term for AI systems that use massive amounts of data and processing power to generate novel output – are reverberating across the technology sector and geopolitics. The use of AI systems to create human-like language, novel images, and even computer code has led to a frenzy of interest from the business and investor community. It has also prompted warnings from policymakers and members of civil society who are concerned about potential misuse of these tools. Here, we explain what generative AI is, describe the key parts of its emerging technology ecosystem, and identify geopolitical and regulatory issues that companies will have to navigate as they develop and deploy this technology.

Understanding generative AI

Generative AI is an industry term for AI algorithms that can create novel outputs – such as images, text, or music – in response to a set of inputs. While generative AI has been the subject of research and experimentation for several years, recent improvements in the size, *This number is poised to grow rapidly as money pours into the sector.* sophistication, and capabilities of generative AI models have thrust the industry into the business and political spotlight.

The growing buzz is best exemplified by ChatGPT, an AI application launched in November by the U.S.-based company OpenAI. ChatGPT is a type of generative AI system known as a large language model (LLM), which is trained on large amounts of text and can perform a variety of tasks that require a nuanced use of language.

While sometimes called a chatbot, ChatGPT and similar systems have a variety of potential uses beyond holding life-like conversations. ChatGPT, for example, can analyze and write computer code in addition to understanding, composing, or summarizing text in multiple languages. Other potentially valuable uses of LLMs and other generative AI systems will be discovered as private sector companies experiment with the technology and its associated applications and business models

ChatGPT's uncanny ability to formulate human-sounding (although not always accurate) responses to user questions has sparked excitement beyond the AI expert community. A publicly available version of ChatGPT has attracted millions of users since its debut in late November. Microsoft, which is an investor in OpenAI and has an exclusive licensing deal with the company, has begun designing ChatGPT functionality into its Bing search engine and other software.

Other companies are vying to develop and commercialize their own generative AI applications. Google recently invested significant funds into an OpenAI rival called Anthropic and has announced plans to incorporate LLMs into its search engine and other products via the chatbot Bard, based on its LaMDA large language model. Chinese technology companies – including Huawei, Alibaba, Baidu, and Tencent – are also developing generative AI platforms and applications.

U.S., European, and Chinese companies have also launched text-to-image models that generate novel images in response to user prompts; text-to-audio models, which can create music in a

variety of styles; and systems that can help scientists discover new protein structures for use in synthetic biology and other life science applications.

One recent study by the research firm CB Insights tallied roughly 250 different generative AI applications currently in the marketplace. This number is poised to grow rapidly as money pours into the sector and companies compete to commercialize generative AI.

Industry excitement is tempered by safety concerns – and global politics

As generative AI models have attracted headlines, policymakers and civil society have begun sounding alarms about the potential risks associated with these systems. These include the potential for generative AI to be used to propagate political disinformation, write malicious code, or cause other harms by generating misleading, inaccurate, harmful, or otherwise inappropriate content.

Some of these concerns have already started to materialize. Microsoft has faced controversy after new chat capabilities it built into its Bing search engine started returning disturbing responses to user input, including stating to a New York Times reporter that it was "tired of being stuck in this chatbox" and likening an Associated Press reporter to Hitler, for example.

Other observers have expressed broader concern that sophisticated generative AI systems may lead to lead to job losses in creative industries or support academic misconduct in schools and universities.

Because LLMs do not understand the concepts they are discussing, they can produce natural-sounding language that contains logical errors or other mistakes, or other output that could be harmful or biased. These issues have led some industry skeptics to worry that a premature rush to commercialize the technology – with companies competing to launch and commercialize LLMs before safety and quality issues have been resolved or effective safeguards have been put in place – could lead to a regulatory backlash that harms the AI sector over the long run.

Further complicating matters, the generative AI revolution is unfolding in a hotly charged geopolitical environment, marked by escalating nation-state competition over leadership in advanced computing, AI, semiconductors, data, and other important technology inputs. As governments around the world take an interest in generative AI and its potential impacts on economic and national security, the risk of disruptive new policies, regulations, or other government actions that could redound on AI developers and the broader technology ecosystem is increasing.

U.S. and Chinese companies will continue to dominate the emerging ecosystem

It will take time to determine the true size of the commercial opportunity presented by generative AI, and to determine which companies and sectors will eventually capture most of the economic value from it. However, major building blocks of the generative AI ecosystem are already coming into focus. They include:

- **Hardware, including advanced semiconductors.** This especially involves graphics processing units (GPUs) and other application-specific integrated circuits (ASICs) that have been specially designed to handle AI workloads. Semiconductors provide the computing power, or compute, required to train generative AI systems. Computing power is also important for inference, where an AI system uses a pre-trained algorithm to make predictions based on a particular set of inputs. As models grow larger and more sophisticated, access to semiconductors that can crunch massive amounts of data as efficiently as possible will be important for advancing the cutting edge of generative AI.

- **Data.** Large data sets are another key ingredient of generative AI. Data can be either proprietary, open-source, or a combination. ChatGPT, for example, was trained on a number of open-source repositories of text scraped from Wikipedia and other written sources. Other generative AI systems may use narrower, domain-specific data sets, such as scientific journals or databases of protein shapes, depending on their intended function.

- **Cloud computing infrastructure.** Cloud-based services are widely used by developers of LLMs and other generative AI systems for training and inference, and to host applications. Along with deploying their own specialized semiconductors optimized for handling AI workloads, some large U.S. cloud players have been investing in companies that are developing generative AI systems or building generative AI systems themselves.

- **Foundation models.** Foundation models are a term applied to more general-purpose AI models that underpin ChatGPT and other user-facing generative AI applications. Foundation models can be adapted and fine-tuned for use in a variety of specific user-facing applications. ChatGPT, for example, represents a refinement of an OpenAI foundation model known as GPT-3.5, in which the AI model is fine-tuned in part using input and feedback from human trainers to produce output that is better aligned with human preferences. Along with developers' access to increasing amount of data and compute, new approaches to training large language models based on limited amounts of human input have been critical for powering recent breakthroughs in the technology.

- **End-user applications.** AI applications that are ultimately used by consumers or businesses can be built directly by cloud providers or specialist AI developers, as in the case of OpenAI and ChatGPT. They may also be built by third parties that license AI models and package them into user-facing products. AI developers can offer companies working on user-facing applications access to their models through application programming interfaces, or APIs.

To date, most of the generative AI systems responsible for the recent wave of interest in the sector have been launched by U.S.-based firms, while U.S. cloud providers have emerged as key players, along with chipmakers, in generative AI's underlying tech ecosystem. The ecosystem will continue to evolve as companies and investors experiment with commercialization of the technology and as regulators deliberate about how or whether to put guardrails in place.

Although OpenAI and other developers have attracted large amounts of funding at lofty valuations, it is still open to debate which companies (and by extension, countries) capture most

of the value from generative AI and are able to best capitalize on its benefits. Along with the evolving uses and risks associated with generative AI, the evolving structure of the industry will influence how governments around the world view the sector. Governments' policy and regulatory responses to generative AI will further influence how the technology and its supporting ecosystem evolve.

Generative AI has geopolitical and regulatory implications

The generative AI revolution is unfolding in political environment marked by escalating strategic technology competition between the U.S. and China; efforts by the European Commission, the Biden administration, and governments in other advanced and emerging economies to curb the influence of large technology platforms and to encourage domestic tech sector development; and concerns about the privacy, fairness, and safety of AI systems. Below we lay out some of the key geopolitical, policy, and regulatory issues that companies will have to navigate as they attempt to develop and commercialize generative AI.

Generative AI challenges China's evolving AI regulatory regime

ChatGPT has garnered significant attention in China as Chinese users have flooded the service with requests in Mandarin. ChatGPT's facility with the Chinese language and its ability to mimic well-known writers has sparked excitement about the potential commercial opportunities that could be open to Chinese technology firms that produce generative AI models optimized for the domestic market.

Companies will need to show that their innovative new products and services do not undermine the political status quo.

China's consumer-focused domestic tech firms will likely have an advantage here, given their inherent advantages in understanding the Chinese language and culture. In recent months, nearly every major Chinese technology company has announced plans to release its own version of ChatGPT. Baidu, which deploys China's most popular search engine and is an AI leader in areas such as autonomous vehicles, said it would finish testing an advanced LLM and ChatGPT competitor, known as Ernie Bot, as early as March 2023. Ernie Bot 3.0 has already been benchmarked against ChatGPT3 and Baidu claims it outperforms the OpenAI version on most standard metrics. Alibaba, JD, NetEase, and iFlytek have also stated their intentions to develop and integrate LLMs into their products and services.

Given the regulatory focus on both tech platforms and AI algorithms over the past year by a range of Chinese government bodies, big tech platforms are likely to step carefully as they roll out chatbots or other generative AI tools. However, the competitive environment in this space is already intense and likely to become even more so in the coming months. Baidu's investment in its Ernie platform appears to be aimed at wrestling market share back from other major players like Bytedance and Tencent. Chinese players will likely attempt to focus their LLM application deployment in enterprise verticals such as cloud and autonomous vehicles, and potentially the metaverse. Interestingly, there is no comparable counterpart in China to OpenAI, which is a smaller and nimbler player and does not have the same political exposure to issues such as disinformation as larger U.S. internet platforms

How quickly large Chinese technology firms are able to deploy generative AI applications will depend on several factors, including the Chinese government's stance on the technology and the

impact of U.S. export controls restricting Chinese companies' access to advanced semiconductors.

Early indications suggest that Beijing will take a cautious approach to generative AI, with large-language models likely to receive special scrutiny. In late February, Chinese regulators warned large Chinese tech firms not to offer ChatGPT-like services, while informing Chinese companies that they should check in with regulators before launching similar systems. OpenAI has since restricted access to the app in the country, citing "legal and regulatory considerations." In the long run, Beijing is unlikely to tolerate an advanced U.S. AI system operating in the country when almost all major U.S.-based technology platforms, including search engines, are banned.

Chinese startups and established tech companies will have to carefully balance their desire to capitalize on a major new commercial opportunity with the need to comply with Beijing's political and regulatory agenda. Companies will need to show that their innovative new products and services do not undermine the political status quo.

AI developers will likely take precautions to try to ensure that chatbots or text-to-image models do not reflect political viewpoints that could be perceived as undermining the authority of the Communist Party or its leadership. However, fine-tuning applications to ensure that they do not produce controversial content is likely to be difficult and time-consuming and will likely require significant trial-and-error.

Chinese tech platforms in the AI algorithm development space are on notice that regulators ready to intensity scrutiny of their software, particularly if any major controversies are revealed as algorithms become widely used. China's main internet regulator the Cyberspace Administration of China (CAC) has issued regulations on AI algorithms and established a registry. Although full details of how CAC will use the registry remain unclear, Chinese regulators appear to be far ahead of counterparts in the U.S. and EU in putting in place mechanisms for reviewing AI algorithms and achieving higher levels of transparency about how they function. CAC regulations focus more on specific types of algorithms used in e-commerce, such as recommendation algorithms, and the role they play in information dissemination, to avoid endangering national security or the public interest. Chatbots are likely to fall into this category, for example, if Baidu's Ernie chatbot is deployed with the firm's Apollo platform for interactions with drivers or passengers in autonomous vehicles, which could impact public safety.

How Beijing responds to errant generative AI models – such as a chatbot that generates politically incorrect responses to user prompts – will be an important signpost for the future of the industry in China. A harsh government reaction could lead to a chilling effect on innovation and experimentation. Beijing may also encourage companies to focus on developing industrial applications of LLMs, such as generating computer code or new drug candidates, over more consumer-facing applications of the technology.

Hardware critical to LLMs is a battleground in U.S.-China technology competition

The geopolitics of semiconductors will also influence the trajectory of generative AI in China. Chinese companies will be developing LLMs in an environment where U.S. restrictions limit access to some types of advanced semiconductors, including Nvidia's A100 and H100 series of GPUs. ChatGPT was trained on roughly 10,000 Nvidia A100 GPUs, and future LLMs will have even greater compute requirements. A detailed look at U.S. export controls, which contain specific thresholds for GPU-to-GPU communications, suggests they were structured to limit access to

GPU features that are particularly useful for training LLMs and other compute-intensive AI applications.

If Chinese companies are unable to easily access cutting-edge AI chips, over time they will struggle to develop LLMs with the same level of efficiency as Western firms. This could make Chinese generative AI products less competitive in the global marketplace. China's leading technology companies have stockpiled some advanced semiconductors, including A100s, and flexibility in hardware platform configurations will still allow for training of LLMs. For example, Baidu's ERNIE 3.0 Titan app is trained on Nvidia V100 GPUs and Huawei's AI-optimized semiconductors, which are not currently subject to U.S. export controls. However, as LLMs grow in size and sophistication, it may be harder for Chinese companies to find workarounds. Leading domestic GPU developers such as Biren are also facing constraints on using global foundry leader TSMC to manufacture their advanced chip designs. TSMC has suspended cooperation with Biren and is reviewing whether the firm's latest GPU designs exceed thresholds established in the Commerce Department's October 7 export control package.

Generative AI will fuel European concerns about digital sovereignty

Large language models have burst into the political and popular mainstream at a critical time for EU tech policy, as the 27-member bloc gears up to enforce its new rulebook for large technology platforms and continues to hammer out details of the EU's AI Act. Underlying both of these efforts is a concern that the EU has become overly reliant on foreign technology companies that may not necessarily share European values.

One key question is the role that large U.S. – and to a lesser extent, Chinese – cloud service providers will ultimately play in the generative AI ecosystem. While well-funded startups have led the recent wave of innovation that is sparking interest in generative AI, the biggest U.S. cloud companies could have major advantages in commercialization of the technology through access to massive amounts of data and compute, their ability to pay competitive salaries to leading programming talent, and their commercial relationships with a wide variety of companies in all sectors of the global economy.

If large U.S. cloud players become the main way that most companies access the next wave of AI innovation, it would present a conundrum for Brussels: Being able tap into these capabilities via the cloud may make it easier for companies that lack expertise in generative AI to incorporate the technology into their businesses, ultimately accelerating AI uptake and making Europe more economically competitive. However, this would also exacerbate concerns about the influence of large U.S. tech platforms and provide political ammunition to more protectionist politicians in Brussels and some member states who favor policies that would promote digital sovereignty and the growth of more European technology champions.

The growing interest in generative AI is contributing to political pressure to tweak the EU's AI Act to include rules for the governance of general-purpose AI systems, a category that could cover LLMs and text-to-image models but was not subject to specific requirements in the European Commission's original AI Act proposal. Even before ChatGPT started making headlines, debate was underway in the European Parliament about how more general-purpose language models should be regulated, if at all, under the EU's risk-based framework, which is focused on "high-risk" user-facing applications. It is highly likely that the final version of the AI Act will include provisions for regulating general-purpose AI systems, including requirements for information-

sharing between developers of general-purpose systems and companies that are using them in specific user-facing applications.

U.S. regulatory efforts will continue to lag other advanced economies

The Biden administration shares some of the EU's concerns about the power of large technology platform companies and has been pushing its own regulatory strategy for AI, including the development of a new, voluntary AI Risk Management Framework. On the domestic regulatory front, the U.S. will likely continue its approach of asking federal agencies with jurisdiction over certain sectors or particular AI uses to issue guidance and rules for how AI systems should be regulated.

The Federal Trade Commission (FTC) and Food and Drug Administration (FDA) are two examples of regulators who will likely end up grappling with issues presented by generative AI applications. Any use of ChatGPT or similar chatbots in medical settings would likely be captured by the FDA's existing framework governing software as a medical device, which is currently being extended to cover issues posed by AI applications. The FTC, which recently set up a new Office of Technology to help manage an increasing technology-related workload, has been stepping up communication on AI issues, including warning companies about potential bias in AI training data. It will likely monitor how generative AI is being deployed to determine whether there are issues presented by these models that could require additional dialogue with industry, tweaks to existing guidance, or new rulemaking. As with other attempts to write detailed rules to cover emerging technologies, this process is likely to be slow and subject to major political wrangling, leading the U.S. to lag other advanced economies in regulating generative AI.

Congress is also taking a growing interest. In January, Representative Ted Lieu (D-CA), one of only a handful of members of Congress with a computer science degree, introduced a resolution directing Congress to scrutinize AI models, noting that the text of the resolution was written by ChatGPT. Lieu has also proposed creating a non-partisan commission to recommend regulations on AI. However, partisan infighting, a lack of prioritization of domestic tech issues among lawmakers, and a preference for light-touch regulation mean that U.S. state legislatures are a more likely source of new laws that could affect the sector than Congress.

In addition to several state data privacy laws that will come into effect in 2023 that could influence how companies deploy AI systems, some state governments have begun to scrutinize generative AI. A state lawmaker in Massachusetts, for example, recently filed a bill drafted by ChatGPT that would require developers of LLMs to register with the state attorney general's office and disclose information about their algorithms. The bill would require companies to undertake risk assessments and implement security measures where necessary. While it is unclear whether the bill will pass, it highlights how state lawmakers are concerned about generative AI models and their ability to cause online and offline harms.

One area where generative AI could generate action at the federal level is around export controls and other restrictions on access to U.S. technology. Concern that generative AI systems could fuel disinformation campaigns, help malicious cyber actors design new hacking tools, or lead to new research breakthroughs that could have military applications is likely to increase pressure from China hawks in the Biden administration and Congress to crack down further on China's access to semiconductors, AI training data sets and other inputs important for generative AI. An upcoming Biden administration executive order on outbound investment and an expected

executive order on data, both of which are likely to contain new measures targeting China and other "adversary" countries, will be important indicators of how generative AI is feeding into these debates.

Finally, the emergence of ChatGPT could complicate the debate in the U.S. over Section 230 of the 1996 Communications Decency Act, which protects intermediaries from legal liability for third-party generated content that is hosted on their platforms. It is unclear whether ChatGPT and other LLMs themselves would qualify for Section 230 protections if they do fall in the category of information content providers (ICPs), though their integration into other products and services – for instance search – may qualify.

Copyright, attribution, and liability issues merit watching

The ability of generative AI systems to render lifelike text or create new images depends on training them on vast quantities of information that may be subject to copyright or governed by creative commons licenses that require attribution when using the material. How or whether generative AI systems can comply with these rules is not yet clear, and will likely be the subject of court cases. There are also unresolved questions about who – if anyone – should be able to assert copyright over the images and text that generative AI systems create. Depending on how courts in different jurisdictions rule, it could create barriers for the commercialization of generative AI. Changing concepts of authorship and attribution introduced by generative AI models may also create calls for new legislation.

Access to data and compute

Beyond China's specific challenges related to U.S. technology restrictions, other governments and companies developing and using the technology will have to grapple with issues related to access to data and compute as generative AI systems continue to grow and evolve. In Europe, pressure to keep up with the U.S. and China in generative AI will add urgency to EU efforts to encourage companies to pool and share data – such as the sector-specific "data spaces" that Brussels is promoting under its recent Data Governance Act. Initiatives such as the U.S. National AI Research Resource, an initiative that aims to ensure that academic institutions and civil society can access the compute and other resources needed to pursue research in AI, will also take on new salience.

As with other technologies that require large amounts of computing power, the significant costs of developing LLMs and related applications will likely favor well-funded and capable technology companies with the financial and human resources to develop and maintain the relevant ecosystems. Providing the energy needed to train large generative AI models is likely to be a particular bottleneck, and how companies address this challenge will shape the evolution of the sector. Just as decisions about where to invest in cryptocurrency mining operations and hyperscale data centers have been influenced in part by access to cheap, abundant, and reliable power, investments in generative AI infrastructure may follow a similar pattern. This could particularly be the case for investments required for training, which need not be located close to end-users. It may be harder to perform this kind of arbitrage when it comes to inference, where the latency, or delay between an AI system receiving an input and returning a result back to a user, may need to be minimized in certain applications.

Finally, as governments compete to lure investment in generative AI, concerns about the technology's energy requirements are likely to have other political side effects. LLMs will

increase urgency around addressing infrastructure issues in countries like India, where access to reliable supplies of water and power has been an issue for the tech sector. It could also give countries with lower power costs advantages in attracting investments. At the same time, concerns about the energy requirements of large models and their related carbon emissions will lead to additional political scrutiny and potentially pressure from investors concerned about the environmental impacts of these technologies.

About ASG

Albright Stonebridge Group (ASG), part of Dentons Global Advisors, is the premier global strategy and commercial diplomacy firm. We help clients understand and successfully navigate the intersection of public, private, and social sectors in international markets. ASG's worldwide team has served clients in more than 120 countries.

The Technology Policy and Strategy Group brings together leading experts in technology, policy, and corporate strategy to help clients navigate complex, high-stakes issues at the intersection of technology and global affairs. Working together with regional and country experts and a broader network of strategic communications and government affairs professionals, our team provides trusted counsel to help business leaders avoid risks and capture opportunities created by digital disruption.

For questions or to arrange a follow-up conversation, please contact:

Kevin Allison
Vice President
Technology Policy and Strategy Group
kallison@albrightstonebridge.com

Anarkalee Perera
Director
Technology Policy and Strategy Group
aperera@albrightstonebridge.com

Paul Triolo
Senior Vice President
Technology Policy and Strategy Group Lead
ptriolo@albrightstonebridge.com