GEOPOLITICAL DRIVERS OF PERSONAL DATA:

THE FOUR HORSEMEN OF THE DATAPOCALYPSE

Paul Smart, Ming-chin Monique Chu, Kieron O'Hara, Les Carr and Wendy Hall

Suggested Citation: Smart, P. R., Chu, M.-c. M., O'Hara, K., Carr, L., & Hall, W. (2019) *Geopolitical Drivers of Personal Data: The Four Horsemen of the Datapocalypse*. University of Southampton, Southampton, UK.

Address correspondence to ps02v@ecs.soton.ac.uk.

Authors

Paul Smart
University of
Southampton

Ming-chin Monique Chu
University of
Southampton

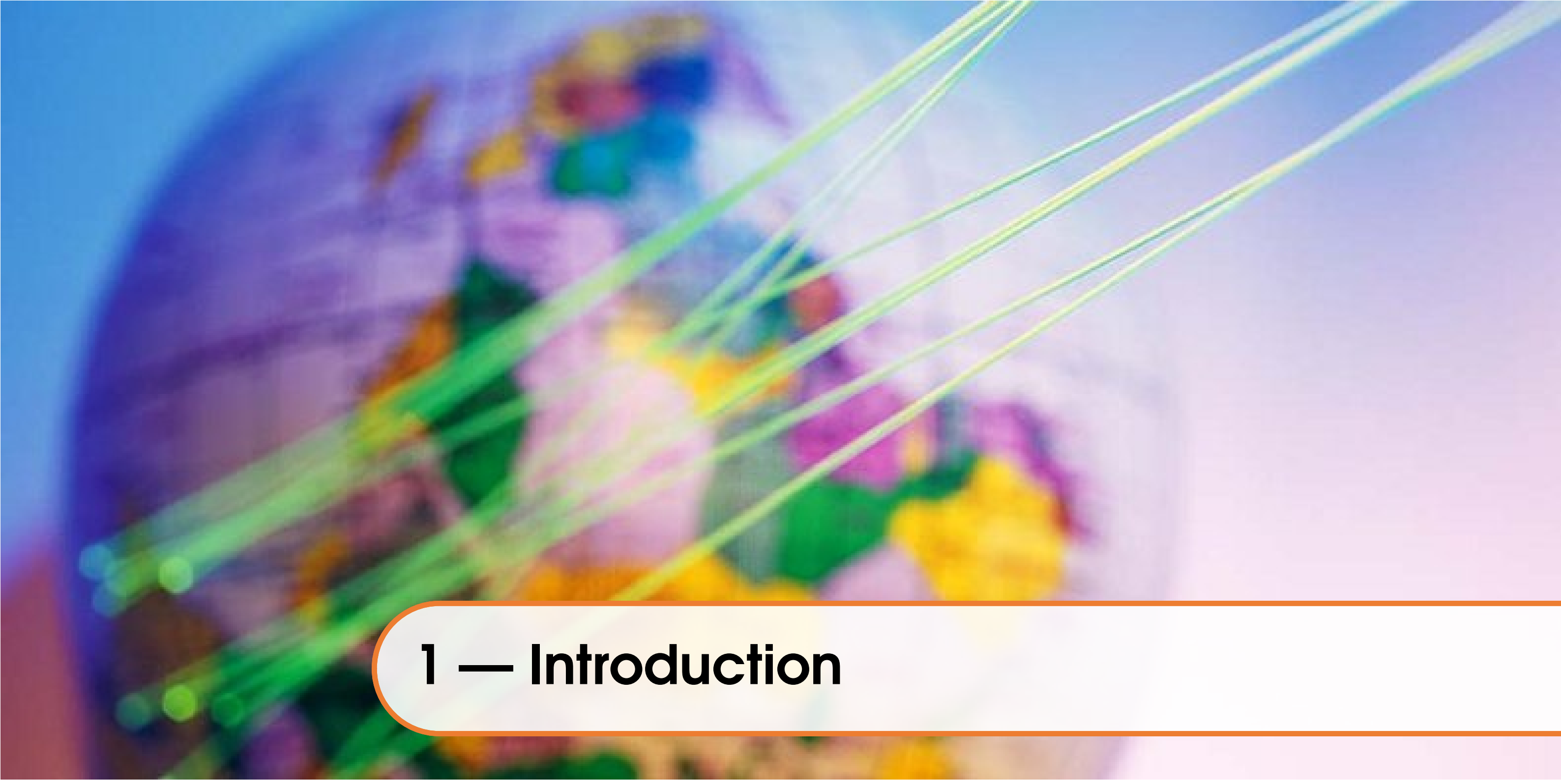
Kieron O'Hara
University of
Southampton

Les Carr
University of
Southampton

Wendy Hall
University of
Southampton

Contents

This page was intentionally left blank.



1 — Introduction

The history of our species is a tale of technological innovation. As new technologies arise (and proliferate), so they alter the kinds of things we can do, both individually and collectively. Our contemporary society is founded on technology. In the absence of technology, society (as we know it) would cease to exist. Technologies are sometimes the catalysts of social change, with innovation heralding the arrival of new social and economic orders (e.g., Landes, 2000). But technologies are not just the catalysts of social change, they are also, at least in some cases, the cornerstones of social reality—the building blocks of the human social world.

The Internet is a signature technology of our time. As with earlier technological innovations, the Internet has exerted a significant influence on society. And, just like other technologies, the Internet has become an intrinsic feature of social life: not just an enabler of social change but also part of the material fabric that realizes social phenomena (Smart, Madaan, & Hall, 2019; Smart & Shadbolt, 2014).

The Internet is, however, a highly 'malleable' technology. As noted by Kieron O'Hara and Wendy Hall (2018):

> The internet is not a monolithic architecture whose existence and form are guaranteed in perpetuity, but a fragile and contingent construction of hardware, software, standards and databases, governed by a wide range of private and public actors whose behaviour is constrained only by voluntary protocols. It is therefore subject to evolution and political pressure. (O'Hara & Hall, 2018, p. 1)

This is important, for the Internet is (almost by definition) a technology with global reach. As the Internet comes into contact with different political and ideological viewpoints, so its form adjusts to reflect those viewpoints. In this sense, the Internet is like a chameleon crawling across a political map of the world. As the chameleon progresses, so its colour changes to reflect the patchwork of political and ideological constituencies that colour the geopolitical map. Taken to the extreme, this shape-shifting capability threatens to overturn the traditional image of the Internet as a global information space with the potential to connect all humanity. In its place, we have the image of a fragmented Internet—an online world subject to ethnolinguistic and geopolitical balkanization.

Relative to the libertarian ideals of the Internet's creators, this image of a fragmented Internet represents something of an apocalyptic scenario. For much of its history, the Internet has been guided by a set of ideals that emphasize the importance of common standards and the free flow of information. These ideals have proved tremendously successful. They have seen the Internet morph into a global technology whose impact can be felt in almost every aspect of human social life. But such success has brought the Internet into the geopolitical spotlight. For some governments, the Internet is seen as a threat. Every form of technologically-mediated social change brings with it the demise of the status quo, and thus it is no surprise that a technology like the Internet should provoke both hope and fear, probably in equal measure. In an effort to contain the threat, and perhaps even tame the beast for a domestic purpose, governments have begun to direct their policy-making efforts to the online realm. In some cases, the policies are intended to protect citizens from so-called 'online harms'. (The UK government's *Online Harms White Paper* [Government of the United Kingdom, 2019] exemplifies such concerns.) In other cases, it is much less clear whose interests are being protected. In the case of China, for example, cyber governance is probably intended to protect the interests of the Chinese Communist Party (CCP) as much as it is the interests of individual Chinese citizens.

As noted by O'Hara and Hall (2018), geopolitical differences in cyber governance pose a significant threat to the status of the Internet as a single, monolithic entity, underpinned by a common set of technological protocols, regulatory conventions, and human aspirations. In place of a single Open Internet, we are witnessing the emergence of a multiplicity of Internet models, each of which is championed by a particular set of actors on the geopolitical stage. The result is what O'Hara and Hall (2018) call the "Four Internets"—a vision in which the erstwhile dominance of the Open Internet is challenged by the arrival of three newcomers: the Commercial, Bourgeois, and Paternal Internets.

The Four Internets arise as a result of the policy decisions countenanced by national and regional governments, but such decisions occur against a wider geopolitical backdrop—one that speaks to the tensions and rivalries that exist between nation states. This is significant, because the Internet is not just a technology that supports the exchange of information, it is also a technology that enables further forms of technological innovation. Of particular importance is the way in which the Internet provides opportunities for advances in Artificial Intelligence (AI). While the United States (US) undoubtedly leads the world in AI research and development, its continued dominance in this area is by no means assured. China, for example, is widely seen to have something of a data advantage relative to the US, and this may have implications for China's ambitions to become a global AI superpower. (As *The Economist* recently put it, China is "the Saudi Arabia of data" [The Economist, 2017a].) The result is a concern about the Internet and its implications for global power struggles. This concern is rooted in the status of the Internet as a form of data-gathering machinery and the subsequent implications this has for AI innovation. Data is, of course, independently valuable, in the sense that it provides opportunities for insight, decision-making, and effective action. But data is especially important when it comes to AI research and development (Hall & Pesenti, 2017). This stems from the fact that contemporary AI systems are driven by 'data-hungry' machine learning systems. Just as oil fuelled the Industrial Age, so data is widely regarded as the fuel that will drive the AI era. And if the Internet is the means by which data can be acquired, then nations that are able to harness the data-gathering capabilities of the Internet are likely to enjoy a strategic advantage in the battle for AI supremacy. Here we see the importance of regulatory efforts that affect the Internet's data-gathering capabilities. While Western nations worry about the threat such capabilities pose to individual privacy, China is actively seeking to extend these capabilities. Its plans for a so-called Social Credit System (SCS), for example, situate data collection, data aggregation, and data processing at the very heart of its plans for social governance.

In the race for AI supremacy, however, data is not the only thing that matters. Data may be the oil that drives the AI revolution, but power and 'progress' (if we care to call it that) cannot be obtained by fuel alone; we still need engines to do useful work. The quest for AI supremacy is ultimately a quest for capabilities—it is the attempt to build machines that are able to achieve certain things. Data may be relevant to the acquisition (and exercise) of those capabilities, but it is only one of the ingredients on the list. A useful parallel, perhaps, is with the attempt to build an aircraft that can travel at supersonic speeds. Access to oil—and highly refined oil, at that—is clearly important if the project is ever to get off the ground. But considerable effort must also be spent designing a jet engine. Absent this exercise in advanced engineering—which requires both human expertise and access to state-of-the-art manufacturing facilities—then a lofty ambition is likely to remain in a permanently grounded state.

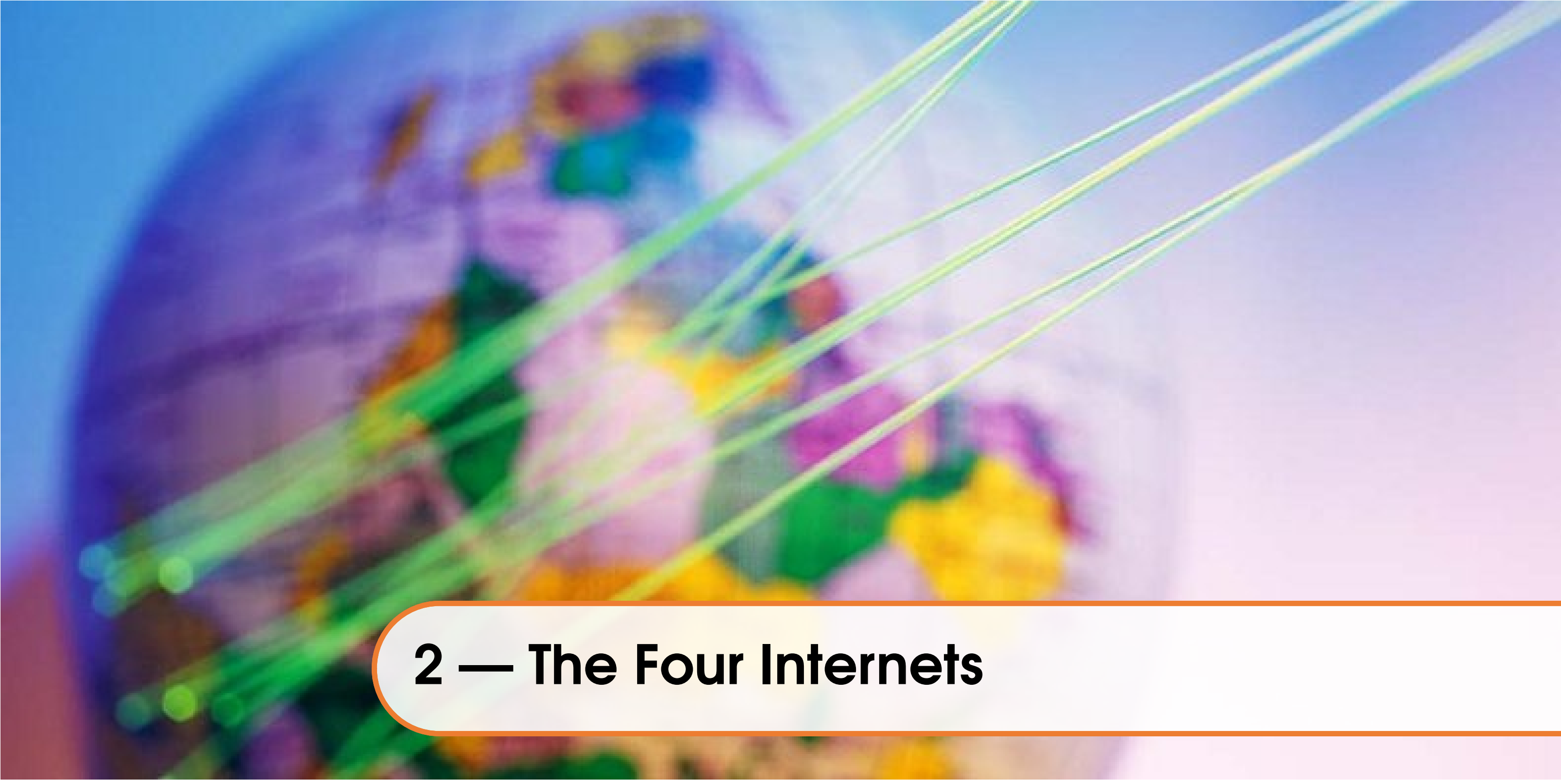
When it comes to AI, then, we need to consider how a nation's AI ambitions may be constrained by more than data. Here we see the importance of global shifts in the semiconductor industry—the industry that provides AI systems with the computer chips they need to consume and digest vast quantities of data. While China is widely seen to have an advantage when it comes to data, it is less clear that it has the upper-hand when it comes to the design and manufacture of high-performance chips. This, then, serves as a potential constraint on China's AI ambitions. It also complicates an assessment of China's capacity to press maximal benefit from the data-gathering capabilities of its domestic Internet. China has long expressed a desire to reduce its dependence on foreign semiconductor technology, but despite an accelerated drive to promote self-sufficiency, China was still left importing more than $300bn of computer chips last year. . . more, in fact, than it spent importing oil (see Lucas, 2019).

A consideration of the global semiconductor industry is thus important if we are to understand the complex mix of forces and factors that are shaping government policy. Given the fundamental importance of computer chips to data processing, AI capabilities, and Internet functionality, the dynamics of the semiconductor industry is likely to have important implications for our understanding of the motivations behind government policy, as well as our understanding of the geopolitical consequences of such policies. As noted by Ming-chin Monique Chu (2013), major shifts in the global distribution of design expertise and semiconductor manufacturing capabilities are likely to have significant implications for the balance of global economic and military power.

The present report seeks to survey some of the issues pertaining to Internet evolution, the quest for AI supremacy, and the status of data as a strategic resource. Clearly, given the complexity of such issues, it will not be possible to provide an exhaustive analysis. Instead, the present report attempts to survey a limited part of the relevant terrain. For the sake of simplicity, we focus much of our attention on China. Extensions of the present work could attempt to compare and contrast the approaches to cyber, data, AI, and social governance adopted by other major players on the geopolitical stage, for example, Europe, the US, and Russia.

The structure of the report is as follows: In Section 2, we present the four Internet models described by O'Hara and Hall (2018). Section 3 shifts the focus of attention to AI-related matters. In particular, we discuss the nature of China's AI ambitions and outline some of the factors that speak to these ambitions. Section 4 focuses on China's SCS. We discuss the current status of the SCS and some of the problems associated with its implementation. We also contrast the SCS with conventional financial credit rating systems and online reputation systems. Section 5 focuses on issues of data value and the factors that might inform attempts at data valorization. Finally, Section 6 concludes the report and outlines some areas for further work.

This page was intentionally left blank.



2 — The Four Internets

The Berlin Wall symbolized a world divided and it defined an entire era. Today, remnants of that wall sit inside this museum [the Newseum] where they belong, and the new iconic infrastructure of our age is the Internet. Instead of division, it stands for connection. But even as networks spread to nations around the globe, virtual walls are cropping up in place of visible walls.

—Hillary Clinton (2010)

For much of its history, the design, management, and use of the Internet has been guided by the libertarian ideals of its Silicon Valley creators. The result is what O'Hara and Hall (2018) call the Open Internet—an Internet that supports the free flow of information. This, however, is not the only vision of what the Internet is, or, perhaps better, what it *ought* to be. As the Internet has expanded to become a global phenomenon, it has had to contend with a panoply of different political ideologies, many of which are opposed to the purist libertarian ideals that underpin the Open Internet. This has yielded competing visions of what the Internet ought to be, how it ought to be managed, and whose interests it ought to serve. Such visions undoubtedly challenge the hegemony of the Open Internet, and they may even lead to a fragmentation or splintering of the Internet into geopolitically and (perhaps) technologically distinct systems.

In this section, we outline the four Internet models described by Kieron O'Hara and Wendy Hall (2018). Each of these models reflects a particular view of how the Internet should be run.

2.1 The Open Internet

The Open Internet is characterized by a commitment to the free flow of information, with minimal restrictions on the type of information that can be transmitted. At the heart of this vision lies the notion of *network neutrality*, which mandates the equal treatment of Internet communications, regardless of their content, origin, destination, and purpose.

Historically, the Open Internet was the first to appear. It is particularly associated with the libertarian techies of Silicon Valley in California. For the most part, proponents of the Open Internet favour technological solutions to problems, as opposed to solutions rooted in government intervention. Problems (e.g., fake news) are still recognized in the context of the Open Internet; however, such problems are typically framed as technical problems that can be resolved via technological innovation (as opposed to government intervention). This does not mean that the Open Internet is opposed to regulation, however. Proponents of the Open Internet recognize that some forms of regulation are required to ensure the basic principles upon which the Open Internet is founded. Regulation may thus be required to ensure the free (and impartial) flow of information across the global network.

2.2 The Bourgeois Internet

As the Internet has grown in popularity, so have concerns about its capacity to create problems, both for individual citizens and for society-at-large. The Bourgeois Internet can be seen as a response to some of the problems raised by the Open Internet's commitment to the free flow of information. Of particular concern is the permissiveness of the Open Internet, the fact that it allows for the dissemination of harmful or inappropriate content. In response, the Bourgeois Internet seeks to minimize social harms, while simultaneously preserving the libertarian emphasis on autonomy and freedom. It is a vision that is premised on the idea of everyone behaving well within a set of well-understood social norms and conventions. From a geopolitical perspective, the Bourgeois Internet is exemplified by the policy statements and regulatory actions of the European Union (EU).

The primary purpose of regulation in the context of the Bourgeois Internet is to protect citizens against a number of perceived 'harms'. In essence, the aim is to provide a regulatory framework that addresses some of the perceived shortcomings of the Open Internet model, with a view to creating a safe online environment. Such sentiments are nicely summarized by the opening statements of a recent UK government white paper:

> The government wants the UK to be the safest place in the world to go online, and the best place to start and grow a digital business. Given the prevalence of illegal and harmful content online, and the level of public concern about online harms, not just in the UK but worldwide, we believe that the digital economy urgently needs a new regulatory framework to improve our citizens' safety online. This will rebuild public confidence and set clear expectations of companies, allowing our citizens to enjoy more safely the benefits that online services offer. (Government of the United Kingdom, 2019, p. 5)

The nature of the online harms that motivate a regulatory response are many and varied. Once concern relates to the nature of the information that is available in the online environment. Specific examples include pornographic material or content expressing extreme political or religious views.

In recent years, the epistemic effects of the Internet have also been a cause for concern. The potential for misinformation, for example, has led to worries about 'fake news' and its capacity to sway public opinion. In the worst case, fake news is deemed to pose a threat to democracy, allowing both foreign and domestic actors to influence the outcome of political elections (Persily, 2017). Proponents of an Open Internet model would no doubt favour a technological response to this particular problem; however, it is unclear whether technological advances, by themselves, can provide an effective solution. One of the problems here is that technological advances tend to benefit both sides of a technological arms race. We see this in the case of recent advances in AI technology, which have helped to support the detection of fake news content (Ruchansky, Seo, & Liu, 2017), but which have also complicated the original problem via the emergence of so-called "deep fakes" (Chesney & Citron, forthcoming).

Proponents of the Bourgeois Internet are also sceptical of the capacity of market mechanisms to resolve online harms. (In this respect, they differ from proponents of the Commercial Internet, which is discussed below.) In the EU, for example, there have been calls for greater regulation of Internet companies such as Airbnb, which is accused of accentuating urban inequality, changing the face of neighbourhoods, and swamping beautiful European cities with tourists (Henley, 2019; Hinsliff, 2018).

The EU has also expressed concerns about the dominance of major US technology firms in the European online market. As noted by a 2017 article in *The Economist*, this marks a difference between Europe and America, with the latter expressing greater tolerance of market dominance:

> The big transatlantic gap is in the policing of dominant firms (known in Europe as Article 102 cases, after the relevant passage in the EU treaty). Europe's trustbusters have been far more likely to worry that a dominant company, of the sort that technology industries tend to produce, will force rivals out of business, leaving consumers facing less choice, higher prices and worse services. Trustbusting in America, in contrast, has taken its cue from the economist Joseph Schumpeter who believed that the promise of monopoly profits is a spur to the innovation and risk-taking that drives economic growth. In this view, the dominance of tech firms is likelier to attract competition than to crush it. (The Economist, 2017b)

The extent to which innovation is stymied by the EU's regulatory approach is unclear; however, it may be harder to innovate in the context of the Bourgeois Internet. The introduction of the EU General Data Protection Regulation (GDPR), for example, may limit companies' access to personal data and thus hinder the EU's capacity to compete in a big data economy. Such competition extends to the development of AI products and services, many of which are driven by 'data-hungry' machine learning algorithms.

2.3 The Commercial Internet

A more commercial view of the Internet is characteristic of the US Republicans in Washington DC. This is what is called the Commercial Internet. Proponents of the Commercial Internet often highlight the economic value and potential of the Internet. In particular, the online realm is typically seen as a forum for the operation of market mechanisms. Online resources are cast as a form of private property whose owners can monetize them, exclude others from using them, and seek market rates for their use. According to proponents of the Commercial Internet, decisions about information flow ought to be left to private actors. If, for example, someone wanted to restrict the flow of information out of their domain, that is their 'business'. As with the Open and Bourgeois Internets, proponents of the Commercial Internet recognize the existence of problems; however, (free) market mechanisms are often seen as a means of resolving such problems.

Although we have cast the Commercial Internet as a compatriot of the Open Internet, there are clear differences between the models. One difference emerges in respect of the notion of net neutrality. As we have seen, net neutrality is a signature feature of the Open Internet; however, it is typically opposed by proponents of the Commercial Internet. Perhaps unsurprisingly, commercial companies (specifically, Internet providers) are opposed to net neutrality on the grounds that it restricts their freedom to manage their networks as they see fit. From a commercial standpoint, this risks undermining the extent to which companies can maximize their profits. A strict adherence to net neutrality, for example, would prohibit a company's ability to give preferential treatment to some forms of network communication in exchange for elevated fees.

Opposition to net neutrality (and thus the Open Internet) is also rooted in ideological differences. While proponents of the Open Internet may acknowledge the need for regulation to safeguard (e.g.) net neutrality, neoliberal, free-market thinkers are often opposed to government interference in market processes. Inasmuch as the Internet is deemed to constitute a form of private property—as something that can be owned—then the proponent of the Commercial

Internet does not see why commercial Internet organizations should be prevented from exploiting their property for the purpose of making a profit.

O'Hara and Hall (2018) note that this property-centric vision threatens to undermine the original vision of the Internet as a universal information space. They cite the example of social networking sites that bypass interoperability mechanisms. These, they suggest, lead to the emergence of platform-specific data silos, which are embedded within the larger technological fabric of the Internet but inaccessible to external systems. Imagine, for the sake of argument, that the Web comes to be dominated by a select group of social media systems, each of which is closed to external systems. Relative to such a state-of-affairs, it would be difficult to regard the Web (or Internet) as a global information space. Instead, what we have is something akin to a "splinternet"—a fragmentation of the Internet into a collection of independent systems, each of which services the interests of its own community and focuses its attention on proprietary, platform-specific forms of technological innovation.

2.4 The Paternal Internet

The Paternal Internet is typified by the regulatory practices and policy positions of China. According to this model, the Internet is viewed as subservient to the interests of a particular social group. In contrast to the Open Internet, the Paternal Internet is rooted in attempts to control the flow of information, and any information deemed incompatible with the values of a ruling elite is typically censored.

Supporters of the Paternal Internet see the Internet's surveillant tendencies as an opportunity to steer, nudge, or even coerce a population into performing or refraining from particular actions. In this sense, the Internet is sometimes seen as a tool for social governance.

As with the Bourgeois Internet, proponents of the Paternal Internet seek to exert control over the online realm via the development of regulatory frameworks. Government intervention is, however, more robust in the case of the Paternal Internet. In addition, the regulation often serves a different purpose. While the regulatory objectives of the Bourgeois Internet are directed to safeguarding citizens and preserving the public good, the purpose of regulation in the case of the Paternal Internet is often directed towards a particular group of individuals. In the case of China, for example, cyber governance is just as likely to be driven by the need to ensure the continued survival and dominance of the CCP as it is to protect individual citizens from online harm (see King, Pan, & Roberts, 2013).

Proponents of the Paternal Internet typically seek to limit exposure to external influences and shape the Internet according to a domestic political agenda. China epitomizes this approach via the imposition of robust censorship schemes, which limit access to information that lies beyond its geopolitical borders. Partly as a result of such policies, China has limited the extent to which Western technology companies, such as Google and Facebook, are able to penetrate the Chinese market. The upshot is that China's Internet is dominated by a limited number of home-grown (i.e., Chinese) commercial giants, e.g., Baidu, Alibaba, and Tencent. Such commercial giants have arisen as a result of market forces, but also as a result of China's (protectionist) policies. China has, in short, fostered the emergence of a limited number of private companies that now dominate its domestic Internet. In contrast to its Bourgeois cousin, the Paternal Internet is not necessarily opposed to such forms of monopolization. While market monopolization by private companies is seen as a potential threat to competition by the EU; in China, such forms of monopolization may serve as an acceptable substitute for public ownership. Given that is arguably easier to control 3–4 large Internet companies than it is to control a multitude of small ones, China's protectionist policies can be seen to work in concert with its authoritarian approach to cyber governance. In essence, China has cultivated an online environment that is not just compatible with authoritarianism but also more amenable to government control.

China's Internet has also been shaped in other ways. Consider, for example, China's response to the fake news phenomenon mentioned earlier (see Section 2.2). While Western countries have yet to develop a coordinated response to this particular epistemic threat, China has sought to tackle the problem by resorting to legislation. China's punitive stance towards 'fake' news is epitomized by the so-called "retweet rule"—if false or misleading information is propagated (retweeted) more than 500 times, then the originator is subject to custodial penalties (see Creemers, 2018b). The introduction of this rule increases the risk associated with certain kinds of online systems. Systems where individual posts can be accessed by anyone and propagated indiscriminately are particularly hazardous, since it is difficult for individuals to judge the popularity (and thus the subsequent distribution) of individual postings. The result of the retweet rule, according to Creemers (2018b), is a shift in the social media landscape, from open systems like Weibo to more private messaging systems like WeChat. While the ostensibly private nature of such systems might be seen to count against the interests of the Chinese government (we might assume the government would prefer completely open systems for monitoring purposes), systems like WeChat do afford greater levels of government control over the online information environment. In particular, targeted messaging systems reduce the rate of information dissemination within a social network and thus limit the spread of 'false' or 'harmful' information. As an added bonus, a reduction in the rate of information distribution allows extra time for government authorities to evaluate and forestall the spread of subversive or politically-sensitive content (Creemers, 2017).

Such examples help us understand how the Chinese government shapes the online ecology in a way that supports its social control objectives. They also help us appreciate the bidirectional influences between the government and the Internet—the Internet informs government policy, while government policy influences the nature of the online ecology.

The Paternal Internet poses a significant threat to the dominance of the Open Internet. As noted by Segal (2018), the Internet has long been an American project, but the rise of China as a global cyber-superpower provides it with opportunities to reshape "cyberspace in its own image." According to Segal:

> If this happens, the Internet will be less global and less open. A major part of it will run Chinese applications over Chinese-made hardware. And Beijing will reap the economic, diplomatic, national security, and intelligence benefits that once flowed to Washington. (Segal, 2018)

2.5 The Moscow Mule Spoiler Model

In addition to the four Internets outlined above, it is possible to identify another vision of the Internet, one that is rooted in subversion. This is a vision dubbed the Moscow Mule Spoiler Model.

This particular vision doesn't constitute a fifth Internet, because it doesn't suggest a particular form, function, or architecture to build. Rather, the aim is to subvert the functionality of an existing Internet model—to turn the power of an existing Internet model against itself. When it comes to the Open Internet, for example, the emphasis on openness and the free flow of information provides opportunities for malign intervention, such as the spread of fake news and malware. Note that the aim here is not to destroy or eradicate an existing Internet model—while proponents of the Moscow Mule Spoiler Model are happy to troll the Internet, they still require a functioning Internet to troll, and thus they have no incentive to undermine it totally.

Although many governments are no doubt guilty of exploiting the Internet for malign purposes, O'Hara and Hall (2018) locate the main epicentre of thinking in this space in Russia, where President Vladimir Putin's brand of mystical nationalism, inspired by the philosopher Ivan

Ilyin, mandates aggressive manipulation of the information space of its perceived enemies.[1]

While the Moscow Mule Spoiler Model does not qualify as a fifth Internet, it nevertheless plays a crucial role in supporting the emergence of alternatives to the Open Internet. In particular, by exposing some of the inherent vulnerabilities of the Open Internet, it provides the basis for more authoritarian approaches to cyber governance. In this sense, the Moscow Mule Spoiler Model is not so much a separate Internet as it is a catalyst that drives the process of Internet evolution and its speciation into a multiplicity of different forms. Note, for example, that Russian cyberattacks on the Internets of both the US and Europe highlight shortcomings in the Open Internet that then feed into the narratives associated with more paternalistic or authoritarian approaches to cyber governance. The need for alternative models is thus legitimated as a result of the apparent threat the Open Internet poses to social stability and democracy. Such shortcomings also highlight the need for strong central leadership in protecting individuals from the hazards of the online realm.

2.6 Comparing the Four Internets

One way of understanding the similarities and differences between the four Internet models discussed above is by directing our attention to the notion of ‘openness’, which is a central element of the Open Internet model. In an Information and Communication Technology (ICT) context, openness is commonly conceptualized with respect to the following:

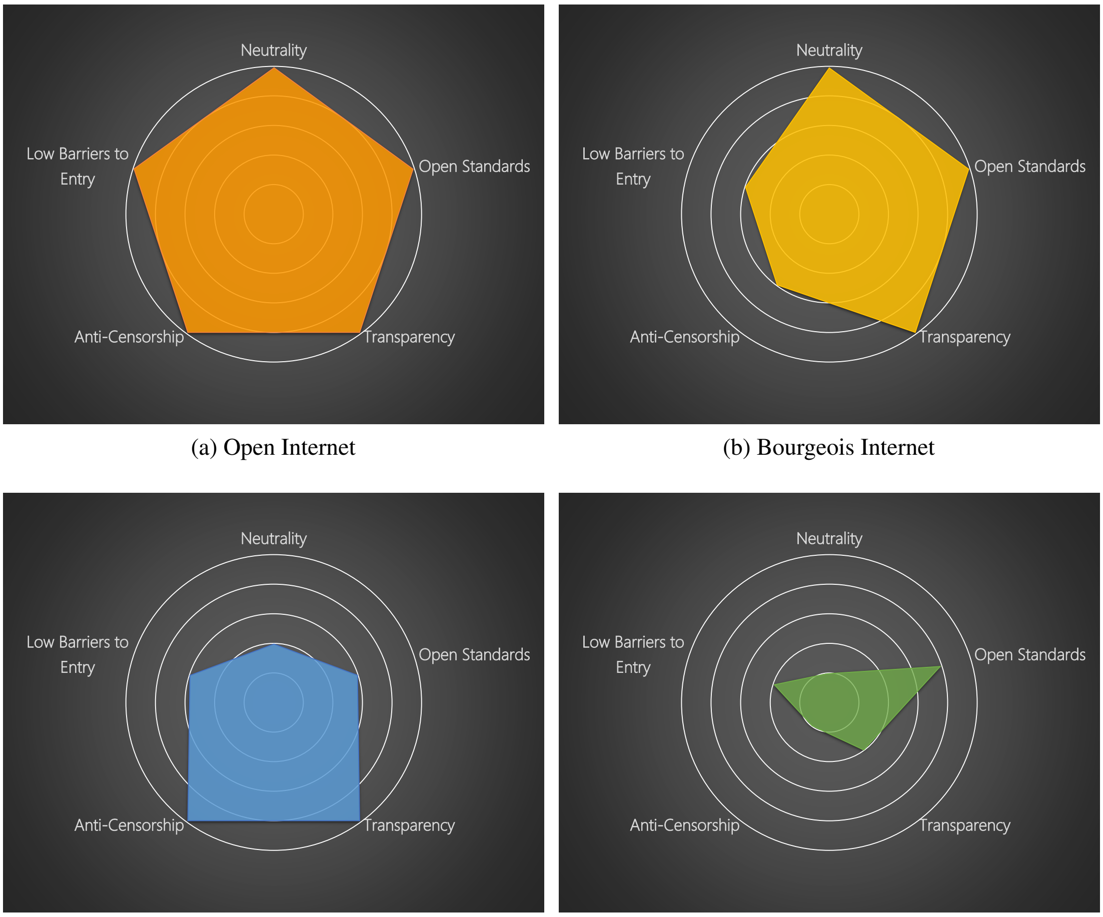
- **Net Neutrality:** The Open Internet is characterized by a commitment to the free flow of information, with minimal restrictions on the type of information that can be transmitted. At the heart of this vision lies the notion of network neutrality, which mandates the equal treatment of information, regardless of its content, origin, destination, and use.
- **Open Standards:** Endorses the use of publicly-available and royalty-free standards.
- **Transparency:** No interference with communicated content.
- **Anti-Censorship:** No suppression or control over what can be published or transmitted.
- **Low Barriers to Entry:** Minimal costs imposed on new entrants.

The Open Internet can be seen to embody a maximal form of openness relative to these features. It thus endorses maximal forms of transparency, net neutrality, and so on. Other Internet models will tend to score less highly on each of these dimensions. Some insight into the relative differences and similarities between the aforementioned Internet models can thus be gained by charting the position of each Internet model with respect to openness-related features. Figure 2.1 exemplifies this approach to comparing Internet models.

In addition to issues of openness, there are a number of other ways in which Internet models might be compared. These include the nature of the approach adopted to resolve problems and the types of social actors whose interests are protected by cyber governance policies. The following list summarizes these additional distinctions:

- **Solution Strategy:** How are problems (e.g., fake news) typically resolved? Possible responses are market solutions, regulatory solutions, and technological solutions.
- **Market Monopolization:** How are large, dominant Internet companies perceived within each of the models? Possible responses are positive, negative, and neutral.

[1] China has also been accused of launching cyberattacks against not-military targets. According to Western intelligence reports, China’s military operates a specialized unit, called P.L.A. Unit 61398, which, it is claimed, is the source of a number of prominent attacks against government and civilian targets outside mainland China (Sanger, Barboza, & Perlroth, 2013).



(a) Open Internet

(b) Bourgeois Internet

(c) Commercial Internet

(d) Paternal Internet

Figure 2.1: Multi-dimensional characterization of the four Internets. [Neutrality: All Internet communications treated equally; Open Standards: Publicly-available and royalty-free standards; Transparency: No interference with communicated content; Anti-Censorship: No suppression or control over what can be published or transmitted; Low Barriers to Entry: Minimal costs imposed on new entrants.]

- **Social Actors:** Whose interests are best served by cyber governance policies. Possible responses are the global community, government bodies, nation states, corporations, and individuals.

Table 2.1 illustrates how the four Internet models compare with respect to these features.

2.7 Internet Evolution and the Geopolitical Niche

"Every success contains the seeds of its own overthrow," writes the British journalist and businessman, Matt Ridley (2003). "Every hegemony comes to an end." Ridley's focus is the process of biological evolution, but is there any reason to think that the evolutionary history of the Internet will be any different? The early Internet embodied the libertarian values of its creators, and the resultant model—the Open Internet—has proved to be remarkably successful, helping to establish the Internet as a global phenomenon. Such success, however, comes with an attendant risk: The Internet now inhabits a complex geopolitical landscape that is marked by the ceaseless struggles between different (and often competing) ideologies. As the Internet adapts

Table 2.1: Comparison of Internet models.

Dimension	Open	Commercial	Bourgeois	Paternal
Solution Strategy	Technological	Market	Regulatory	Regulatory/ Technological
Market Monopolization	Neutral	Positive	Negative	Positive
Social Actors	Global Community	Corporations	Individuals	Government Bodies/Nation States

to this geopolitical niche, it may undergo a transformation, similar to the process of adaptive radiation, yielding a multiplicity of distinct forms. Our characterization of the Four Internets marks an attempt to understand some of these forms. However, these are not the only forms that might emerge. Neither is the survival of *any* form guaranteed in the longer term. The fate of any species is unclear, and the Internet is no different. The Internet may have many possible futures. . . including none at all.

3 — Artificial Intelligence

Artificial intelligence is the future, not only for Russia, but for all humankind. It comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world.

—Vladimir Putin (2017)

In March 2016, Lee Seedol—one of South Korea's top professional Go players—was pitted against a computer system developed by Google DeepMind, a UK company specializing in deep machine learning. To the surprise of many, the computer system—dubbed AlphaGo—achieved a stunning victory, beating its human opponent 4–1 in a five-game match. In one sense, the victory was unremarkable. Coming almost two decades after IBM's computer, Deep Blue, beat world chess champion Garry Kasparov, it would be easy to dismiss AlphaGo as merely the latest demonstration of AI's relative superiority (and thus humanity's relative inferiority) in a limited range of game-playing scenarios. There are, however, reasons to think that AlphaGo's triumph marked a seminal moment in the history of AI. Firstly, the success of AlphaGo demonstrated the power and potential of a relatively recent approach to the development of AI systems, namely, the use of deep learning algorithms (Goodfellow, Bengio, & Courville, 2016; LeCun, Bengio, & Hinton, 2015). Such approaches are now at the forefront of AI research, and they are largely responsible for much of the recent interest in AI capabilities. To some extent, there is nothing particularly novel about such approaches. From an architectural perspective, today's deep neural networks resemble their forebears from the 1980s, and although there has been a genuine and important shift in training techniques—a shift that is marked by the use of generative methods (Hinton, 2010)—such techniques are rooted in an intellectual lineage that dates back to the work of Hermann von Helmholtz in the 19th century. This does not mean that the current enthusiasm for deep learning systems is unfounded, for there have been significant breakthroughs in AI research, and the capabilities of today's deep learning systems far outstrips that of earlier AI systems. Such successes are undoubtedly rooted in a number of architectural and algorithmic novelties, but it is also important to appreciate the role of other factors. Such factors include the widespread availability of high-performance computer processors, advances in computer networking and distributed computing, and, finally, access to large-scale bodies of data that can be used for training purposes.

There is a second reason why AlphaGo's success in the game of Go was significant. While Go resembles chess in the sense that its rules can be formally specified, the complexity of Go far exceeds that of chess. After the first two moves of a Chess game, for example, there are 400 possible next moves. In Go, there are close to 130,000. The search space for Go is thus inordinately larger than that seen in the case of chess. According to Silver and Hassabis (2016), two of the researchers behind the development of AlphaGo, "The search space in Go is vast…a number greater than there are atoms in the universe!"

Due to this complexity, many had believed that the development of an AI system that could master the game of Go was still years into the future. And yet here we are. The future, it seems, has arrived somewhat earlier than we expected, and with it has come a complex mix of concerns about the economic, political, and social impacts of AI technology.

As with other countries, China is grappling with these concerns, and its interest in AI no doubt predates Lee Seedol's defeat to AlphaGo. Nevertheless, AlphaGo's success was perhaps keenly felt by China. According to Shotwell (2008), the game of Go has strong historical links with China that go back more than 2,500 years. In this sense, AlphaGo's success may have been perceived as a major wake-up call by the Chinese leadership (see Mozur, 2017). In the aftermath of Lee Seedol's defeat, China proceeded to declare AI as a national priority. As noted by Ding (2018):

> For China, AlphaGo may have demonstrated that advances in AI are linked to national prestige and the perceived status of great powers. Additionally, the types of high-level seminars conducted after AlphaGo indicate that some Chinese policymakers interpreted AlphaGo's victory as having significant implications for military affairs.

In view of this, it is perhaps no surprise that China has sought to expand its capabilities and develop a domestic AI industry. The present section examines the nature of China's ambitions in this area. It also explores some of the factors that are apt to constrain its ambitions. As noted above, much of the recent progress in AI—especially that associated with research into deep learning—has been driven by the availability of high-performance chips and access to copious quantities of training data. China is widely seen to have a strategic advantage when it comes to data (see Section 3.4.1). Although it is by no means clear that access to personal data (or, at any rate, human-generated data) is a prerequisite for *all* forms of AI innovation (see Section 3.8), personal data remains an important resource for both China and the wider world. In this respect, China's AI ambitions are likely to be useful in helping us understand its approach to data, cyber, and social governance.

3.1 Strategic Objectives

In July 2017, China's State Council issued the *New Generation Artificial Intelligence Development Plan* (State Council, 2017). This document—along with its *Made in China 2025* plan (State Council, 2015), released in May 2015—form the core of China's AI strategy.[2]

As outlined in the *New Generation Artificial Intelligence Development Plan*, China has a specific set of AI-related goals in the 2030 timeframe. These goals are summarized in Table 3.1.

3.2 China's View of the AI Landscape

Chinese AI experts and decision-makers are acutely aware of the AI strategies and capabilities of other countries, especially the US, the EU, Japan, and the UK. In a chapter titled "Top-Level Plans," scholars from Tencent's Research Institute and the China Academy of Information and

[2] China is, of course, not alone in developing an AI strategy; similar strategies have been outlined by other national and regional governments (see Dutton, 2018).

Table 3.1: China's AI ambitions (State Council, 2017).

Area	**2020**	**2025**	**2030**
Technology	Important progress and "iconic advances" in a new generation of AI theories and technologies.	Breakthroughs in autonomous learning abilities for AI systems.	Breakthroughs in brain-inspired intelligence. Important impact on international AI research.
Industry	China will have established initial AI technology standards, service systems, and industrial ecological system chains.	Widespread use of AI technology throughout industry.	AI use expanded into social governance and national defence.
Economy	Core industry: 150 billion RMB Related industry: 1 trillion RMB	Core: 400 billion RMB Related: 5 trillion RMB	Core: 1 trillion RMB Related: 10 trillion RMB
Environment	Opening up new applications. Establishing AI ethical norms, policies, and regulations in some areas.	Initial establishment of AI laws and regulations, ethical norms and policy systems, and the formation of AI security assessment and control capabilities.	The construction of more comprehensive AI laws and regulations, and an ethical norms and policy system.

Communications Technology (CAICT) laid out their view of the current international strategic landscape for AI development as follows (Ding, 2018, p. 11–12):

- **Defend the lead America:** "In sum, the United States is, at this point, the country that has introduced the most strategies and policy reports on artificial intelligence strategies. The United States is undoubtedly the forerunner in the field of artificial intelligence research and its every move necessarily affects the fate of all of humanity."
- **Ambitious EU:** "In 2013, the European Union proposed a 10-year Human Brain Project, currently the most important human brain research project in the world."
- **Robot superpower Japan:** "For the past 30 years, Japan has been called the robot superpower and has the world's largest number of robot users, robotics equipment, and service manufacturers."
- **Unwilling to fall behind Britain:** "The UK considers itself to be a global leader in ethical standards for robotics and AI systems. At the same time, this leadership in this area could extend to the field of artificial intelligence regulation."

China's position within this landscape is described by the report's authors as follows:

> In terms of AI, China followed the United States and Canada in releasing a national AI strategy. In the wave of AI industry, our country should go from system follower and move towards being a leader, actively seizing the strategic high ground. (Ding, 2018, p. 12)

3.3 Realizing the Vision

3.3.1 AI Research

The *New Generation Artificial Intelligence Development Plan* calls for progress in a bewildering array of research areas, including (but not limited to) big data intelligence, cross-media intelligence, swarm intelligence, human–machine symbiosis, autonomous intelligent systems, autonomous learning, brain-inspired AI, and quantum-accelerated machine learning. While such a list attests to the scale of China's ambitions in this area, the *New Generation Artificial Intelligence Development Plan* is not, by itself, particularly informative as to where China's priorities lie. Inasmuch as the plan reflects a genuine commitment by the Chinese government to fund all areas of AI research, then there is a danger that it may fail to achieve critical mass in any area. It is also important to note that many of the research areas mentioned in the *New Generation Artificial Intelligence Development Plan* are well-served by existing research efforts outside of mainland China. In this sense, it is difficult to assess the extent to which China's AI research program is truly innovative.

In order to build up an indigenous innovation capacity, China has been constantly raising its research and development budget. In terms of overall expenditure, it is surpassed only by the US. In February 2017, China's *Artificial Intelligence 2.0* plan received "megaproject" designation, which comes with substantial state funding (Ding, 2018).

Apart from central government support, Chinese municipalities have announced plans to boost AI development in local areas. These include the development of a 13.8 billion yuan ($2.12 billion) AI technology park in Beijing's western suburbs (Tan, 2018). According to the official Xinhua news agency, the park will house up to 400 enterprises and have an estimated annual output of 50 billion yuan (cited in Cadell, 2018).

3.3.2 Research Outputs

Given the high levels of investment being directed toward AI research in China, we would expect China to be a major player when it comes to AI research. There is some evidence to support this idea. As early as 2014, for example, China was beginning to surpass the US in terms of the number of journal articles mentioning the terms "deep learning" or "deep neural network." This was noted in the White House's strategic plan for AI research, published in October 2016 (National Science and Technology Council, 2016).

While China may have a growing number of research publications, this does not mean that it is ahead of the game when it comes to *innovative* AI research. There is, however, some evidence to suggest that China is making progress on this front. Data on presentations at the Association for the Advancement of Artificial Intelligence (AAAI) annual conference, widely recognized as a leading AI conference, revealed that Chinese researchers account for roughly 20% of the findings presented, second only to those from the US (see Ding, 2018) (see Figure 3.1). China, however, lags behind both the US and UK in terms of citation impact. According to a McKinsey Global Institute report, research from the US and UK is more influential by citation impact, as measured by the H-index (Barton, Woetzel, Seong, & Tian, 2017).

3.4 Strengths

3.4.1 China's Data Advantage

Access to data is typically regarded as one of China's key strengths in an AI context (e.g., Ding, 2018). This is due to the way in which many AI systems are now developed. For the most part, AI systems are developed using machine learning techniques, and the success of such techniques is often tied to the properties of the data that drives the learning process. Kai Lee (founder of Sinovation Ventures, a venture capital firm) notes that:

> Many practical AI advances are more about having a large amount of continually refreshed data and good-enough AI researchers who can make use of that data, rather than some brilliant AI theoretician who doesn't have as much data. (cited in Larson, 2018)

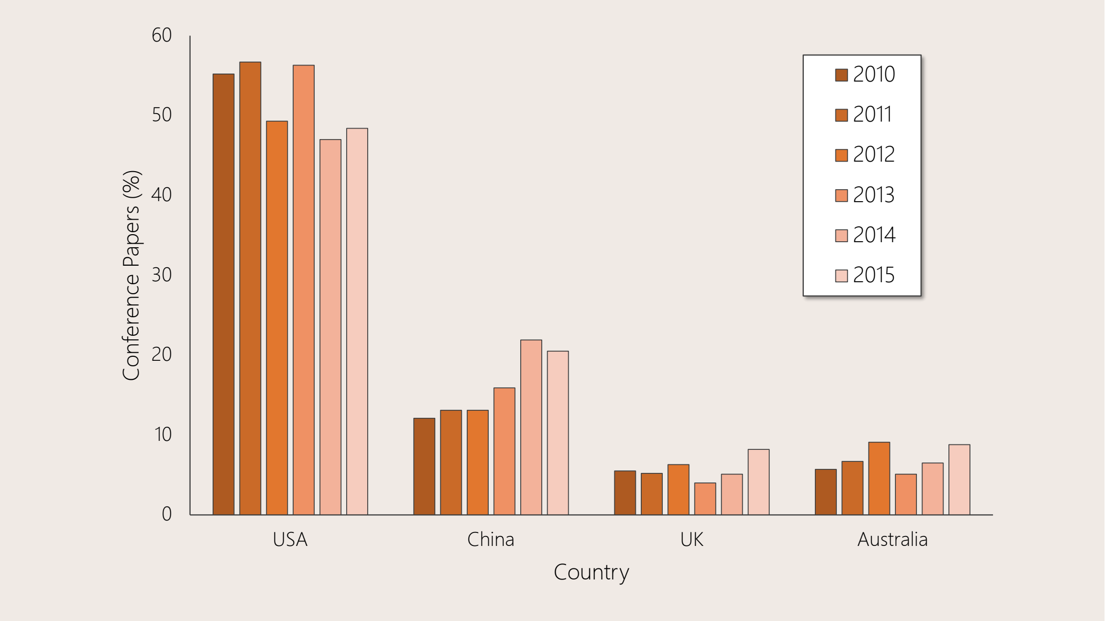


Figure 3.1: AAAI conference presentations by country (source: Ding, 2018).

In general, then, China is typically seen to be ahead of the game when it comes to data-driven forms of intelligence. Courtesy of the abundance of data made available via the Internet, Chinese companies are seen to have an advantage when it comes to the development of AI capabilities. Let us refer to this as the notion of *data advantage*. According to this notion, there is a positive, although no doubt probabilistic, relationship between data quantity and AI capabilities, such that more data equals more intelligence.

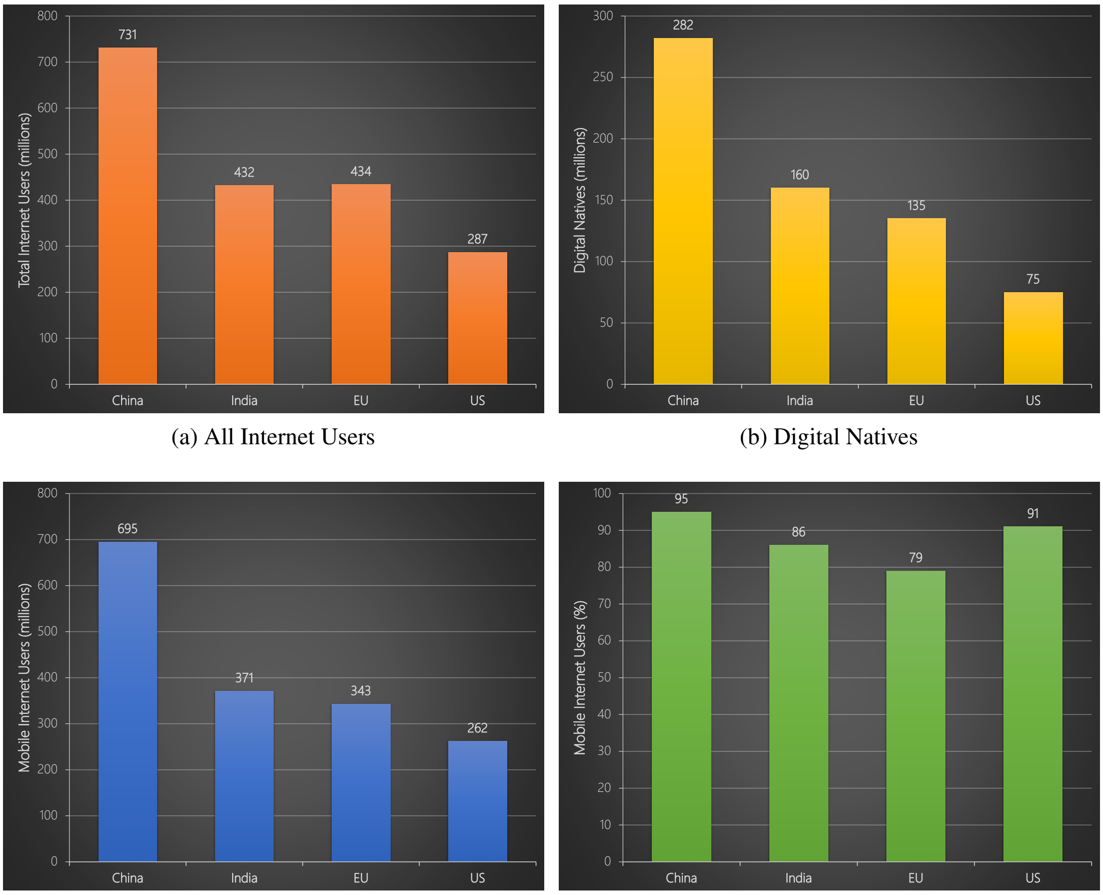
Assuming that the notion of data advantage is correct, then China may be suitably placed to realize its AI ambitions. Indeed, China has been described as "the Saudi Arabia of data" (The Economist, 2017a). A vibrant e-commerce environment, the widespread use of smartphone technology, and the popularity of social media all contribute to an abundance of data, and much of this data is available to companies with AI research divisions. According to Bai Chunli, President of the Chinese Academy of Sciences, China will generate 44 trillion gigabytes of data by 2020, an estimated 20% of global data output (see Ding, 2018). This figure is projected to rise to 30% by 2030 (see Ding, 2018).

3.4.2 Internet Penetration

Internet penetration is one of the factors driving China's growth in domestic data. China has a large number of Internet users who generate data as a by-product of their participation in social media and e-commerce applications. If we define Internet penetration in terms of the number of Internet users, then China's level of Internet penetration compares favourably to many other countries. As of December 2016, for example, China had 731 million Internet users compared to 287 million in the US. It also had 695 million mobile Internet users (compared to 262 million in the US) and 282 million digital natives (compared to 75 million in the US) (Woetzel et al., 2017, p. 5) (see Figure 3.2).

3.4.3 Social Entrenchment

Another reason to think that China has the upper-hand when it comes to data access relates to the notion of social entrenchment. This emphasizes the way in which the Internet is embedded in a



(a) All Internet Users

(b) Digital Natives

(c) Mobile Internet Users

(d) Proportion of Mobile Internet Users

Figure 3.2: Internet penetration in China, India, Europe, and the US (source: Woetzel et al., 2017).

broad array of everyday social activities, to the point where such activities may be difficult, if not impossible, to perform without recourse to the Internet.

Social entrenchment appears to be particularly strong in China. As noted by Larson (2018):

> Every time someone enters a search query into Baidu (China's Google), pays a restaurant tab with WeChat wallet, shops on Taobao (China's Amazon), or catches a ride with Didi (China's Uber), among a plethora of possibilities, those user data can be fed back into algorithms to improve their accuracy. A similar phenomenon is happening in the United States, but China now has 751 million people online, and more than 95% of them access the internet using mobile devices, according to the China Internet Network Information Center. (Larson, 2018)

One example of social entrenchment comes in the form of Ant Financial's Smile-to-Pay service, which is based on computer vision technology (Qi & Xiao, 2018). Instead of relying on a special-purpose payment device to effect a payment (e.g., a smartphone or debit card), a user simply stands in front of a vending machine and smiles to complete a payment (see Figure 3.3).

Systems such as Smile-to-Pay highlight the way in which AI capabilities (in this case, facial recognition) and the Internet are being embedded into everyday life. They also intimate at the presence of a positive reinforcement mechanism that serves to drive the development (and

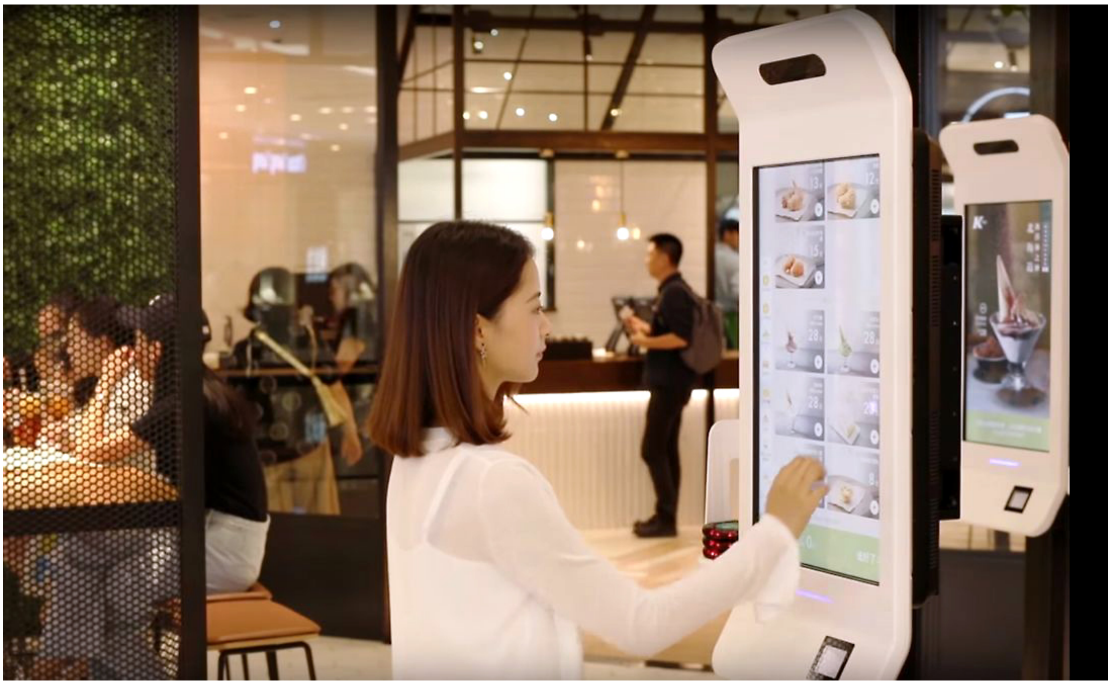


Figure 3.3: Ant Financial's Smile-To-Pay system relies on facial recognition to complete a payment.

perhaps uptake) of AI capabilities. Note, for example, that in the Smile-to-Pay case, the AI system functions not just as the means by which a socially-relevant activity is enabled, it also serves as a system that is nicely poised to collect additional data. This applies not just to the imaging function of the system—the fact that it is able to acquire images of the same face as it appears in different imaging conditions (e.g., different lighting conditions)—it also applies to the capacity of the system to collate various forms of behavioural data about a particular data subject; for example, data pertaining to an individual's economic activities, their daily routines, and their movements around an urban environment.

A second feature of the Smile-to-Pay system is the nature of the behavioural interaction that is encouraged (and perhaps mandated) by the system. Note, for example, that the system does not just require the user to stare blankly into a screen; instead, they are required to smile. Human participation thus involves the adoption of a positive facial expression (a smile), which, according to psychological theories of emotion, most notably, the facial feedback hypothesis (Adelmann & Zajonc, 1989), may translate into a positive emotional state—a genuine feeling of warmth and friendliness. To our mind, this an important, albeit probably overlooked, feature of the Smile-to-Pay system. It is a feature that speaks to issues of technology adoption and the general social acceptance of AI systems as an intrinsic part of social life. Inasmuch as we wish to promote the use of AI systems (for both commercial and social reasons), it is important to consider more than just their capacity to realize some socially-significant function. In addition to this, we ought to consider the *psycho-affective design* of such systems—the way in which the design of a system promotes the adoption of a positive affective stance towards the system. Such a stance may be relevant to issues of acceptance and use, i.e., the extent to which a given user is likely to accept the AI system as a recurring feature of their daily routine. It also, however, speaks to the perceived costs and benefits associated with the use of an AI system.

3.4.4 The Regulatory Environment

China is typically seen to have a relatively lax approach to privacy protection, allowing companies greater scope when it comes to the use (and sometimes sharing) of personal data. As with many other areas of government policy, however, China's approach to personal data is more complicated than might be thought. For a start, data misuse and privacy infringements are a prominent topic of concern for Chinese citizens, and this has prompted the government to respond with new guidelines on the collection, use, and storage of personal data. The 2018 *Personal Information Security Specification*, for example, outlines a number of constraints that apply to companies seeking to collect personal data as part of their commercial activities. While this specification has, on occasion, been regarded as more onerous that the EU GDPR (e.g., Sacks, 2018b), the government's stated position is that the specification aims to provide greater freedom than the GDPR, while simultaneously addressing public concerns about the commercial misuse of personal data (Hong, 2018). In all likelihood, the Chinese government is attempting to strike a balance between the interests of AI companies (vis-à-vis the development of AI technologies) and the interests of the public regarding the protection of personal data. This speaks to an ongoing debate within China regarding the benefits of data liberalization and the need for tighter controls over access to personal data. Commenting on a report authored by Tencent and CAICT researchers, Ding (2018) notes that:

> . . . Tencent and CAICT researchers attribute the success of Silicon Valley to the existence of strong institutions such as copyright and tort law, and they argue that data liberalization is a form of institution building that could spur further innovation. They write, "If there is no government data liberalization policy, many AI applications will become water without a source, a tree without roots." (Ding, 2018)

It is easy to misconstrue this debate as a straightforward choice between AI and privacy. In fact, however, the debate is one that speaks to a more general issue about the best way to realize China's AI ambitions. China's government is acutely aware of the need to promote trust in technology as a means to ensure the acceptance and adoption of AI solutions, and its recent attempts to safeguard personal data are perfectly consistent with this policy.

3.4.5 Foreign Data

In addition to their access to domestically-acquired data, some Chinese companies are able to expand their data portfolios as a result of their commercial footprint in foreign markets. SenseTime, for example, is a major exporter of surveillance technology in markets across Latin America, Africa, and Asia. This provides it with an opportunity to expand its data portfolio beyond the borders of mainland China.

CloudWalk is another AI company that stands to benefit from access to overseas markets. In 2018, CloudWalk signed a deal to provide the government of Zimbabwe with a mass facial recognition system, which aims to monitor all major transportation hubs as well as create a national facial ID database (Feola, 2018; Zaagman, 2018). As noted by Cook (2018), such access is apt to provide opportunities for the refinement of AI capabilities. When it comes to facial recognition, for example, Cook (2018) notes that:

> . . . data and images of ethnic Chinese, Turkic Uighurs, and—under a new deal with Zimbabwe's government—sub-Saharan Africans could collectively enable developers to correct common race-related errors in facial-recognition software and gain market share in other parts of the world. (Cook, 2018)

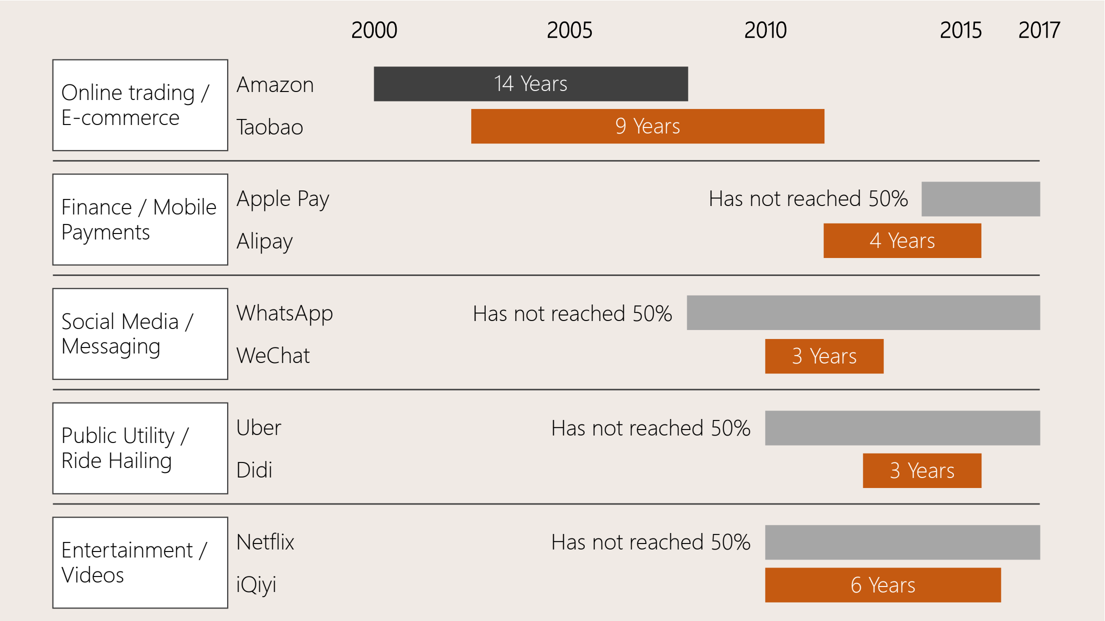


Figure 3.4: The time taken to reach over 50% market penetration by US and Chinese Internet companies (adapted from Lew, 2017).

Here we see the way in which a commercial footprint in overseas markets promises to create something of a positive reinforcement loop, with access to foreign data providing the basis for improvements in AI-related commercial offerings.

The same principle applies to companies whose primary commercial focus is not the AI domain. Huawei's Safe City initiative, for example, uses a combination of surveillance technology and an "integrated command solution" to support policing efforts in the Kenyan capital city of Nairobi (Huawei, 2018). Huawei claims that the regional crime rate has dropped by 46% since the introduction of the system in 2015 (Huawei, 2018).

3.4.6 The Commercial Ecosystem

One of the factors that feeds into the notion of data advantage is China's commercial ecosystem. The key issue here relates to the way in which companies like Baidu, Alibaba, and Tencent have come to dominate China's commercial Internet landscape. From a data access perspective, commercial dominance is important; it means that each company has, in effect, a larger share of the data 'pie', at least relative to situations where a nation's data assets are divided between a multitude of smaller commercial players. (This is especially true when corporate data sharing is prohibited due to competitive interests and/or legislative constraints.) The West does, of course, have its own corporate technology giants, such as Google and Amazon, so we should not assume that market monopolization is a feature unique to the Chinese Internet. Nevertheless, some have suggested that the Chinese Internet is more concentrated than the US. Lew (2017), for instance, notes that a number of Chinese companies have achieved 50 percent market penetration much more rapidly than their US counterparts (see Figure 3.4).

The Chinese government has played an important role in the evolution of China's commercial landscape. There are three aspects to this. The first is China's particular breed of techno-nationalism, which shields domestic companies from foreign competition. By making it difficult for companies such as Facebook and Google to operate in China, the Chinese government has facilitated the emergence of home-grown Internet leviathans such as Tencent and Baidu. This has as much to do with China's political climate as it does with the imposition of government sanctions on foreign companies. China's approach to censorship, for example, is likely to conflict with the values of most Western companies.

State sponsorship has also helped to shape the structure of China's commercial landscape. In 2017, for example, The Ministry of Science and Technology (MOST) recognized the status of companies such as Baidu, Tencent, Alibaba, iFlytek, and SenseTime as national AI champions (see Jing & Dai, 2017). Such endorsements provide these companies with reputational advantages, both nationally and internationally, but they also provide reassurance that such companies will not face competition from state-owned enterprises (see Allen, 2019).

Finally, China has adopted a protectionist policy towards data sharing that limits the extent to which personal data can be stored outside of mainland China. Such a policy makes perfect sense relative to China's AI ambitions. Inasmuch as we accept the claim that data is a valuable resource for AI development, then protectionist policies help to ensure that domestically-acquired data is available for local exploitation.

The upshot of China's approach to data and cyber governance is an Internet landscape that is dominated by a limited number of private companies. From an AI standpoint this is significant, because it provides companies with access to voluminous bodies of domestic data. Beyond mainland China, market monopolization is typically seen as a threat to competition, and governments may take action to prevent such monopolies from forming. Within China, however, monopolization is likely to be seen as somewhat more acceptable. Monopolization may, for example, serve as an acceptable substitute for public ownership. Monopolization also works in concert with China's authoritarian approach to cyber governance. This is because it is easier to monitor the activities of three or four large Internet companies than it is a multitude of small ones.

3.4.7 The Political Context

In addition to issues of data access, it is worth reflecting on the socio-political differences that exist between China and the West. It is easy to overlook such differences in the context of AI-oriented discussions. For the most part, debates about China's AI potential typically highlight the importance of technological innovation. Thus, even when the focus of interest is the notion of data advantage, the broader debate is still one that turns on technological issues, for example, the way in which data access affords opportunities for machine learning. In addition to issues of a technological nature, however, the relative success (or fitness) of an AI solution depends on the environment in which it is deployed. In parallel with the fitness of biological organisms, what matters here is not so much the features of the system itself, but the way in which such features are aligned to the demands of the world in which the relevant system is deployed. There is no real 'success' associated with the discovery of a state-of-the-art AI capability if such a capability is immediately rejected by all those who might be inclined to use it. In this respect, China's commitment to authoritarian forms of government may be of considerable importance. In particular, China's top-down political system enables it to direct and coordinate 'whole-of-society' approaches to the introduction of new technologies. One example of this comes in the form of dockless bike-sharing. Xia and Yang (2018) note that dockless bike-sharing "grew from a novelty concept to a signature aspect of Chinese urban life in less than one year" (p. 41). They go on to note that "In decentralized western systems, dockless bike-sharing typically requires independent approval from various government agencies; support from the Chinese Communist Party allowed China's dockless players to grow quickly" (p. 41). A similar sort of dynamic may emerge in respect in driverless vehicles. That is to say, robust forms of government intervention may be instrumental in supporting the rapid adoption of driverless vehicles, providing China with a first-mover advantage relative to Western countries.[3]

3.5 Weaknesses

While China arguably has a strategic advantage over many other countries in respect of data access, there are a number of areas where China might be seen to lag behind other nations. In comparison to the US, for example, China is commonly seen to be at a disadvantage when it comes to the following (see Allen, 2019):

- **Intellectual Capital.** China still lags behind other countries in terms of the AI talent pool.
- **Software Frameworks.** Common machine learning frameworks are not made in China, and research and development into domestic AI products is mainly based on the work of overseas companies, such as Google's TensorFlow Framework.
- **Semiconductors.** Computer chips are mostly designed in the US and manufactured in Taiwan and South Korea.

China is particularly concerned about its dependence on foreign technology. In the case of the commercial drone company, DJI, for example, all the drone flight software is developed in the US and nearly 35% of the bill of materials for DJI products originates from the US, mostly in the form of computer chips (Allen, 2019, p. 10).

In the interests of developing an independent, domestic AI industry, China has embarked on a number of efforts to address these shortcomings. Here we focus our attention on issues of intellectual capital and the development of AI-oriented software frameworks. China's dependence on semiconductor technology is discussed in Section 3.6.

3.5.1 Intellectual Capital

One of the preconditions for AI innovation is the availability of personnel with sufficient expertise to undertake fundamental research. According to a report by Tsinghua University, by the end of 2017 the international AI talent pool comprised 204,575 individuals, with the US having 28,536 individuals and China in second place with 18,232 (China Institute for Science and Technology Policy, 2018). Such figures may, however, overestimate China's access to the sort of intellectual capital needed to produce ground-breaking research. When one directs their attention to what the Tsinghua report dubs "top AI talent," China falls from second to eighth place, with only 977 individuals compared to 5,518 individuals in the US (reported in Allen, 2019).

Although such figures point to a shortcoming in China's capacity to be a global leader on the research front, it is unclear whether this has any significant bearing on its capacity to apply existing solutions to practical problems. In many cases, the significance of a given AI solution only becomes apparent when it has been deployed and has had sufficient time to exert an effect on social and economic processes. China's strength, in this respect, may lie not so much in its capacity to discover new AI capabilities as its ability to implement AI systems and then deploy them at large scale.

In any case, China is making substantial efforts to bolster its domestic intellectual capital. The *New Generation Artificial Intelligence Development Plan*, for example, mentions a range

[3] There is, of course, an important difference between bikes and cars: bikes are cheap; driverless cars are not. Even with a concerted push from government, the widespread adoption of driverless cars may be out of reach for many in China given the average wealth of Chinese citizens. Developed countries, such as the US, are clearly in a much better position here. A nation's wealth may thus be an important moderator of a government's capacity to promote the widespread adoption of new technologies. While China's government may have the upper-hand, relative to Western countries, in coordinating social responses to AI technology, this advantage may be offset by wealth disparities. In short: China may have a political advantage when it comes to the introduction of socially-disruptive forms of synthetic intelligence, but this political advantage may be offset by an economic shortcoming.

of measures designed to improve AI expertise. These include the introduction of AI as an independent academic discipline and measures to support so-called AI+X (e.g., AI plus biology) professional training.

China also provides a supportive environment for the influx of intellectual capital, as well as other resources. Google, for example, is building an AI centre in Beijing to recruit Chinese talent. As noted by Fischer (2018), this step is all the more remarkable given Google's rather fractious relationship with the Chinese government (Google's search engine has been blocked in China since 2010).

Another factor of potential importance relates to global shifts in migratory policy. Inasmuch as Western democracies impose constraints on immigration, countries with a more open immigration policy are in a position to hoover up global AI talent.

3.5.2 Software Frameworks

AI developers rarely start from scratch when they implement a new AI system. For the most part, AI developers resort to the use of widely available code libraries and, on occasion, cloud-based AI services. Prominent examples of such code libraries, or machine learning frameworks, include the likes of TensorFlow (Google), Spark (Apache), the Microsoft Cognitive Toolkit (Microsoft), and PyTorch (Facebook). None of these frameworks have been developed in China. Chinese official documents often lament this fact and note that much of the research and development of domestic AI products and applications is based around technological platforms that were developed by Western companies.

There are a number of development efforts within China that speak to this issue. Baidu, for example, has launched DuerOS, a system that allows users to embed many AI functionalities, such as natural language processing and image recognition into devices. It has also released a number of open source platforms, such as Apollo for autonomous driving and PaddlePaddle for deep learning (see Zhu, Huang, Chen, & Gao, 2018).

Other companies are also investing in the development of machine learning frameworks. These include SenseTime, a specialist provider of facial recognition and surveillance technology. SenseTime's offering is a proprietary deep learning platform called Parrots, which is optimized for big data processing and distributed computing applications.

Establishing a footprint in the software framework market is important for a number of reasons. In addition to enhancing the reputation and international standing of AI companies, investing in machine learning frameworks also provides companies with an opportunity to influence technical standards, attract AI talent, and shape the nature of the global AI market. Appreciating this fact is important, for it might be thought that the availability of existing machine learning frameworks, such as TensorFlow, is of little consequence to China's strategic AI objectives: If existing frameworks are readily available, then why bother to reinvent the wheel, so to speak, by developing a home-grown alternative? Such a question assumes that the sole purpose of software platforms is to support the development of AI capabilities. In fact, what is closer to the truth is that such platforms function as a form of 'soft power'; they enable companies to influence the shape of the global AI landscape.

3.6 The Semiconductor Industry

3.6.1 Data vs. Chips

As we have seen, data is typically seen as important to AI development, and China is widely regarded to have a strategic advantage when it comes to data access. There is, however, a problem associated with this data-centric vision. The problem is that an emphasis on data risks overlooking the role of computational resources, especially high-performance computer chips, in leveraging the latent value of the big data assets. Processing large datasets is computationally intensive, so access to high-performance chips is necessary if a company is to press maximal benefit from a body of training data. This highlights the value and importance of semiconductor technology vis-à-vis a country's AI ambitions. In short, access to state-of-the-art computer chips may be just as important, if not more important, to a country's AI ambitions than its access to data. At the very least, we might expect that a country's capacity to capitalize on its data advantage is constrained by its ability to secure access to state-of-the-art semiconductor products. The upshot is that the semiconductor industry, and the geopolitics of semiconductor production, are of substantive importance to China's AI ambitions.

China is, of course, aware of this issue.[4] A recent Tsinghua University report, titled "White Paper on AI Chip Technologies," highlights China's awareness of the strategic importance of computer chips for AI applications:

> Whether it is the realization of algorithms, the acquisition [of] a massive database, or the computing capability, the secret behind the rapid development of the AI industry lies in the one and only physical basis, that is, the chips. Therefore, it is no exaggeration to say "No chip, no AI" given the irreplaceable role of [the] AI chip as the cornerstone for AI development and its strategic significance. (Beijing Innovation Center for Future Chips, 2018, p. 2)

3.6.2 Chip Types

From an AI standpoint, we can distinguish between a number of chip types. The most important ones for present purposes are the following:

- **Central Processing Units (CPUs):** These are the primary components of conventional computers. The term "CPU" refers to the electronic circuitry that performs the instructions encoded in a computer program.
- **Graphics Processing Units (GPUs):** In contrast to CPUs, GPUs feature a highly parallel structure that makes them more efficient than CPUs for algorithms that process large amounts of data in parallel. GPUs are typically used for display purposes. They are key components of the graphics cards used in conventional computing devices (e.g., desktop machines).
- **Artificial Intelligence Accelerator Chips (AIACs):** An AIAC is a computer chip that is specifically designed to support AI applications. A popular example of an AIAC is Google's Tensor Processing Unit (TPU), which was designed to run applications developed using Google's TensorFlow framework.
- **Field Programmable Gate Arrays (FPGAs):** FPGAs are integrated circuits that are designed to be configurable. That is to say, FPGAs feature a flexible architecture that can be configured to suit specific purposes. FPGAs are thus said to be "field-programmable" in the sense that they can be programmed by a customer "in the field" to suit specific purposes.

Of these chip types, GPUs and AIACs are particularly important for AI research, especially for research into deep learning. The next two sections provide a brief overview of these chip types.

[4] In fact, China's awareness of the strategic importance of the semiconductor industry has a long history, dating back to the 1950s.

3.6.3 Graphical Processing Units

A consideration of chip types is important due to their differential effectiveness (or differential efficiency) in supporting AI capabilities. When it comes to deep neural networks, for example, CPUs are seldom the best choice for training the neural network—the highly parallelized nature of the computational processing performed by the neural network favours the use of chips with a more parallelized architecture, such as GPUs. The basic point here is that chip architecture plays a crucial role in determining the efficiency of the computational routines that are designed to yield AI capabilities. Such differences can sometimes be significant. Commenting on the role of GPUs in a system designed to recognize conversational speech, Microsoft Technical Fellow, Xuedong Huang, refers to GPUs as "the real weapon" (Cotterell, 2016). The speech system developed by Huang took about a year to complete. Without GPUs, Huang suggests, it would have taken five.

Interestingly, China is often deemed to be particularly dependent on international companies for GPUs. Out of the top 10 US chip-makers, 4 specialize in the design of GPUs, whereas none of the top 10 Chinese chip-making companies specialize in GPUs (Ding, 2018; Fabre, 2018). In fact, most of the world's GPUs are designed by NVIDIA in the US and then manufactured by Taiwan Semiconductor Manufacturing Company (TSMC), which is headquartered in Taiwan (Deloitte, 2017). China by contrast does not have a major manufacturer or designer of GPUs.

The extent to which China's AI ambitions are undermined by this particular shortcoming is unclear. This is because the status of the GPU as the 'chip-of-choice' is being progressively undermined by the shift to more special-purpose chips, such as AIACs and FPGAs. China tends to fare much better in respect of these chip types. According to Fabre (2018), 6 out of the top 10 Chinese chip-makers specialize in AIACs, while 2 specialize in FPGAs. This apparent bias in favour of AIACs is likely to be of particular significance, since chips of this type can sometimes deliver superior performance to GPUs, even when they are manufactured using older process technology. The first generation of Google's TPU, for example, was manufactured using 28 nanometre technology, which is already widely available in China. In 2017, Google claimed that its first generation TPU was 15–30 times faster than GPUs for AI workloads (Condon, 2017). In this respect, China's apparent shortcomings in the design and manufacture of GPUs may be of little consequence to their AI ambitions.

3.6.4 AI Accelerator Chips

As noted above, AIACs can sometimes deliver superior performance for AI applications, as compared to GPUs. This provides a unique opportunity for Chinese companies to expand their market share in the global semiconductor industry. The reason for this is twofold: Firstly, the performance benefits of AIACs do not rely on state-of-the-art manufacturing processes. (As evidenced by Google's first generation TPU, high-performance AIACs can be manufactured using 28 nanometre process technology, which is well within China's domestic chip manufacturing capability.) Secondly, there is a growing global demand for chips that can run special-purpose AI applications, such as speech recognition systems.

China's commercial ecosystem reflects this growing interest in AIACs. Chinese technology giants such as Baidu (in partnership with Intel), Alibaba (via a new subsidiary, Pingtouge), and Huawei (via its HiSilicon subsidiary) have all established semiconductor design divisions focused on the development of AIACs (Allen, 2019). In addition, a number of Chinese AIAC companies have risen to global prominence in recent years. These include Cambricon, which has raised hundreds of millions of dollars in venture capital funding with a multi-billion dollar valuation (Feng, 2018). Cambricon specializes in chips designed to support deep learning applications. Its Cambrian 1A processor (the Cambricon-1A), launched in 2016, is regarded as the world's first commercial deep learning processor for portable computing devices. It featured as part of the

HiSilicon Kiron 970 System-On-Chip (SoC), which was integrated into the Huawei Mate 10 smartphone, Huawei's first AI smartphone, launched in October 2017.

3.6.5 Foreign Investment

One way for China to boost its domestic semiconductor design and manufacturing capabilities is via investments in foreign technology companies. According to a 2017 report by the US President's Council of Advisors on Science and Technology on the semiconductor industry, Chinese companies have been increasingly active in the acquisition space. The report also notes that China often places conditions on access to its market in order to incentivize technology transfer (President's Council of Advisors on Science and Technology, 2017).

Given that China's investment strategy is likely to have important implications for the global semiconductor industry, it is perhaps unsurprising that China's approach to overseas investment has aroused the suspicion of national and regional governments. In the US, China's investment in US chip-makers has come under greater scrutiny by the Committee on Foreign Investment in the United States (CFIUS). In 2017, the White House blocked a state-backed Chinese investment fund from acquiring a US semiconductor company, representing only the fourth time in history that an American president has blocked a corporate acquisition on national security grounds (Donnan, 2017).

As noted by Ding (2018), similar concerns about China's investment strategy have been raised by the EU:

> In his 2017 State of the European Union Speech, Jean-Claude Juncker, president of the EU Commission, rolled out a new framework for screening foreign direct investments into the European Union. The framework identified critical technologies including, artificial intelligence, robotics, semiconductors, technologies with potential dual-use applications, cybersecurity, space or nuclear technology. While Juncker's speech did not explicitly call out Chinese investments, analysts interpreted his warnings about investments from state-owned companies as an implicit reference to China's economic activities. (Ding, 2018, p. 17).

3.6.6 Chips Made in China

A key feature of China's long-term industrial strategy is to reduce its dependence on foreign technology. Historically, China's burgeoning domestic economy in technological devices has been fuelled by chip imports. The introduction of export restrictions on the sale of semiconductor components to China by both the Obama and Trump administrations has undoubtedly exacerbated China's concern about its reliance on chip imports and increased the leadership's resolve to boost its indigenous semiconductor manufacturing capabilities. As noted in a speech by Alibaba co-founder, Jack Ma:

> . . . the market for chips is controlled by Americans. . . And suddenly if they stop selling—what that means, you understand. And that's why China, Japan, and any country, you need core technologies. (cited in Allen, 2019, p. 19)

China has undertaken a number of steps to bolster its domestic semiconductor design and manufacturing capabilities. In 2014, China's, for example, government established a national integrated circuit industry investment fund to reduce its dependence on foreign semiconductors. The initial round of investment was 138.7 billion RMB, which was followed in 2018 by a second round of government funding totalling 300 billion RMB (Allen, 2019).

There is some evidence that this drive towards self-sufficiency is beginning to pay off.

Huawei's Kiron 980, for example, is one of only two smartphone SoCs in the world to use 7 nanometre process technology, the other being Apple's A12 Bionic (S. M. Moore, 2018). Crucially, however, both of these chips were manufactured by TSMC in Taiwan. This highlights an important distinction between the design and manufacturing stages of the semiconductor production process. China has begun to demonstrate significant progress in chip design, but it still lacks a capacity to exploit state-of-the-art manufacturing techniques. As noted by Allen (2019), even the most advanced Chinese semiconductor manufacturers are only able to use 14 nanometre technology, which international firms such as Intel and Samsung achieved in 2014. Allen (2019) goes on to note that China's most advanced semiconductor manufacturer, Semiconductor Manufacturing International Corporation (SMIC), hopes to reach 7 nanometre manufacturing in the early 2020s; however, this would still place China behind other countries. South Korean firm, Samsung, for example, has set its sights on producing 5 nanometre SoCs by 2020 (Moore-Colyer, 2019). TSMC is also moving towards 5 nanometre process technology. Its 5 nanometer manufacturing process is currently described as being at the "risk production" stage, meaning that "the company believes it has finished the process, but initial customers are taking a chance that it will work for their designs" (S. K. Moore, 2019, p. 9).

3.6.7 Chip Imports

Despite its drive towards self-sufficiency, China remains heavily dependent on chip imports. Historically, China has been one of the world's largest consumers of integrated circuits, and much of this consumption has been driven by imports. With the publication of the *Made in China 2025* plan by the Chinese State Council in 2015, it became clear that China aimed to reduce its reliance on imported technology.[5] The *Made in China 2025* plan outlined a number of strategic goals in relation to semiconductor self-sufficiency. In particular, the plan specified a target of 40% self-sufficiency by 2020, growing to 70% by 2025.

At present, it is unclear whether China stands to achieve these targets. According to the 2018 National Imported Key Commodity Value Index, published by the State General Administration of Customs, China imported 417.57 billion integrated circuits in 2018, with a total value of RMB 2058.41 billion ($312 billion), an increase of 19.8% compared to 2017 (Lucas, 2019). This was the first time that China's semiconductor imports exceeded $300 billion.

3.6.8 Taiwan

From a geopolitical standpoint, Taiwan's proximity to mainland China is likely to be a crucial factor in understanding China's prospects in the global semiconductor industry. Taiwan is undoubtedly one of the most important countries in the world when it comes to the production of integrated circuits. Taiwan is home to two of the world's largest semiconductor companies, namely, TSMC and United Microelectronics Corporation (UMC), both of which specialize in the fabrication of integrated circuits. It is important to note that overarching semiconductor production process is one that can be divided into three phases: design, fabrication, and packaging and testing. China has traditionally specialized in the third, and arguably the least valuable of these phases, namely, the packaging and testing phase. Its domestic shortcomings, relative to semiconductor production, thus lie in the areas of design and fabrication. This is supported by the conclusions of a 2016 report by PwC, which places China in third place, behind the US and Taiwan, in respect of both design and fabrication capability (PwC, 2016).

In terms of China's capacity to address the shortcomings, Taiwan is likely to be of crucial importance. Chu (2013), for example, notes that Taiwan is a prominent source of intellectual capital (know-how), managerial expertise, and fabrication technology, all of which feed into

[5] In fact, China has long expressed a desire to reduce its dependence on foreign technology. In this sense, the *Made in China 2025* plan does not mark a radical shift in government policy.

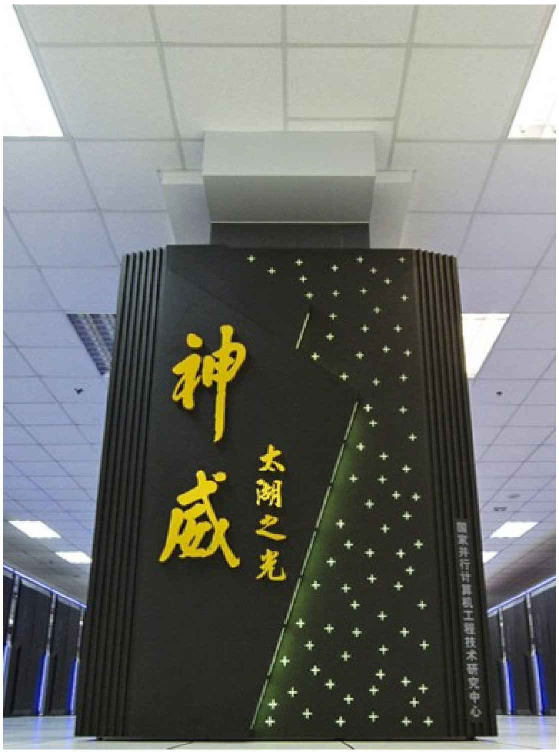


Figure 3.5: The Sunway TaihuLight supercomputer.

China's domestic semiconductor production capabilities. Chu (2016) goes on to note the geopolitical and geo-economic implications of this gradual migration of knowledge, expertise, and technology to mainland China. Many Taiwanese firms have set up their subsidiaries in mainland China for decades, and this 'Silicon Gold Rush' has accelerated since the early 2000s due to the introduction of incentives put in place by the Chinese government. In terms of economic security, emerging semiconductor companies in mainland China may one day become a strong rival of Taiwanese companies, and this may ultimately undermine the status of Taiwan as the epicentre of the global semiconductor industry.

As noted by Chu (2016), such shifts are relevant to contemporary debates about the balance of global power. With the advent of advanced weapon systems, AI-driven capabilities, and the overall digitization of the contemporary battlespace, semiconductors are now a critical element of a country's capacity to project military power and (by implication) exert political influence.

3.7 AI and Supercomputers

3.7.1 Supercomputers

As we have seen, China's capacity to leverage its data advantage is linked to the availability of high-performance computer chips. This claim, however, needs to be treated with some caution, for the primary value of data when it comes to AI capabilities is its capacity to drive machine learning. That is to say, data is valuable insofar as it enables a fledgling AI system to adjust its inner representational and computational economy so as to coordinate its response outputs with respect to statistical regularities that inhere in the target domain of interest. Relative to issues of computational power, it is arguably this phase of system development—the machine learning phase—that presents the greatest challenge to the developers of AI systems. This is especially true when it comes to the effort to evaluate new AI algorithms and progress the current

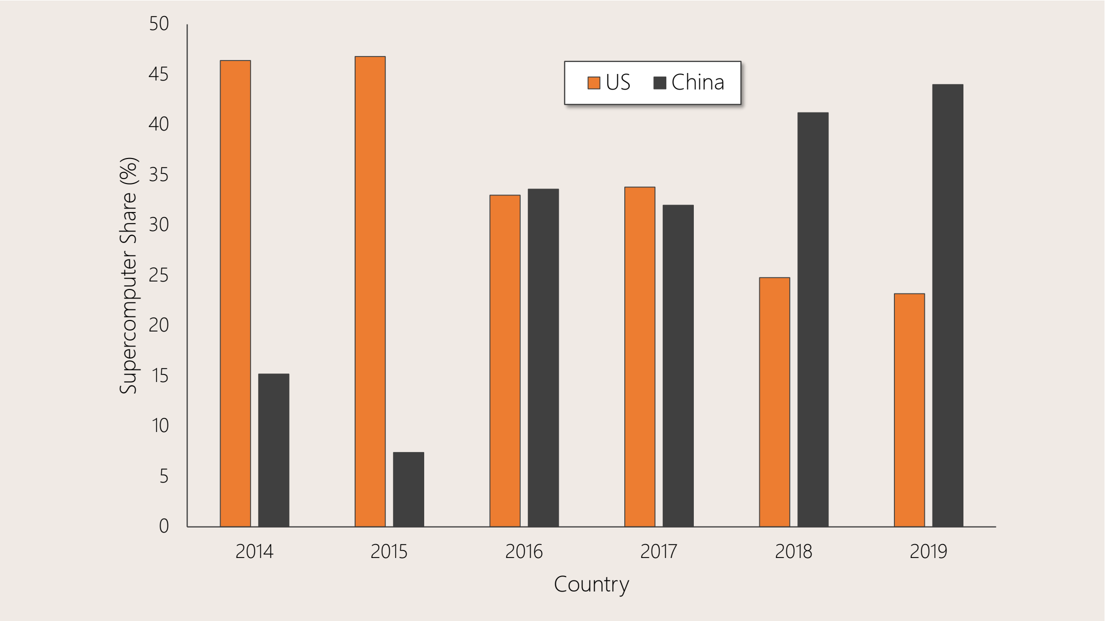


Figure 3.6: Proportion of the top 500 supercomputer facilities owned by the US and China (source: https://www.top500.org).

state-of-the-art.

The importance of computational power for the discovery and development of innovative AI solutions, highlights the potential significance of supercomputing facilities to China's AI ambitions. China has invested heavily in such facilities in recent years. Following the promotion of a National Integrated Circuit Industry in 2014, with an investment of more than $20 billion, China produced one of the worlds' fastest supercomputers, the Sunway TaihuLight (see Figure 3.5), which has a calculation speed of 93 petaflops per second. Apart from the fact that it is, at present, the third fastest computer in the world, the TaihuLight is notable for being built without the use of US processors (Barton et al., 2017, p. 8).

As noted by Lu et al. (2018), China has made substantial progress with regard to the development of supercomputing facilities. According to the Top500 list, a list of the world's highest-performing supercomputer sites, China's global share of supercomputers stood at 76 systems (15.2%) as of 2014, which was a distant second to the US at 232 systems (46.4%). By 2019, however, China had overtaken the US with 44% of the highest-performing supercomputers. The US share, by contrast, fell to 23% (see Figure 3.6).

China is currently planning to boost its national computing power by developing an exascale computer, which it hopes will enter service in 2020. As noted by Chen (2017), the advent of exascale computing represents a milestone in computer engineering:

> An exascale computer is defined as one that can carry out one billion billion calculations per second. It is not only 10 times faster than Sunway Taihulight. . . but equal to the calculation power of all the world's top 500 super computers combined. (Chen, 2017)

It is, as yet, unclear how such advances will affect the shape of the AI research landscape. While computing power is generally seen as a boon to AI research, there are significant costs associated with the use of supercomputing facilities, especially when it comes to electricity consumption.

3.7.2 A Quantum Leap?

From an AI perspective, the importance of computer chips lies in their information processing efficiency—their capacity to perform a required set of computations within a certain timeframe. The advent of quantum computing, however, promises to transform our understanding of computational power. According to Jian-wei Pan, a leading Chinese quantum scientist, “the first general-purpose Chinese quantum computer could have a million times the computing power of all other computers presently in the world” (cited in Lin & Singer, 2017). This is clearly important when it comes to a consideration of the way in which future technological innovation may affect the global balance of AI power. Inasmuch as AI innovation is tied to issues of computing power, then the advent of quantum computing may provide a means for nation states to effectively circumvent the constraints associated with conventional semiconductor technologies.

China is renowned for its achievements in quantum computing. In August 2016, China had its own ‘Sputnik moment’ with the launch of the world’s first quantum satellite (Yu, 2016). Two years later, this satellite, nicknamed Micius, was used to achieve another first, something akin to an ‘Alexander Bell moment’: the world’s first video call made via quantum-encrypted communications (Liao et al., 2018).

The Chinese government is clearly committed to supporting advances in quantum communications and quantum computing. It is, for example, supporting the development of a multi-billion dollar research facility in Hefei, Anhui province, dubbed the National Laboratory for Quantum Information Sciences. Due for completion in 2020, the laboratory is expected to support research into quantum metrology and the development of quantum computers (Zaagman, 2018).

3.8 Data-Driven Intelligence

3.8.1 The Value of Data: Does Size Matter?

What is the relationship between AI capabilities and data availability? Does access to ever-increasing quantities of data inexorably lead to improvements in AI performance, or is the situation much more complicated than this simple proposal would suggest?

From an AI perspective, data is clearly important for at least some purposes. Given the contemporary focus on machine learning systems, especially deep learning systems, much of the value of data (from an AI perspective) inheres in the capacity of some dataset to support the acquisition of some (intelligent) capability. That is to say, data is valuable to the extent that it enables an AI system to configure its internal parameters so as to exhibit successful performance within some target domain of interest.

It is at this point that the simplistic relationship between data quantity and intelligence is called into question. The value of data (*qua* machine learning resource) is not so much the size of the dataset as it is the potential of that dataset to support successful learning. In many cases, of course, the two are related: larger datasets provide more opportunities for learning and thus yield better performance when it comes to the AI system’s real-world operation. But the mere size of a dataset is no guarantee of value. Imagine, for the sake of argument, that we seek to train a deep learning system to recognize images of cats, and we have access to a large, and ostensibly valuable, training resource—a repository containing a billion-plus cat-related images. At first sight, it is difficult to see how such a resource could fail to make a positive contribution to the learning process. If, however, we discover, that the repository contains a billion-plus copies of a single image, then our judgements of value are apt to be revised: a vast repository of identical images is no more useful for machine learning than is a repository containing a single image. We can, after all, present a single image to the same machine learning system multiple times. The point of this example is to remind us that while size is often a good indicator of value, it is by no means the only thing that matters.

3.8.2 The Right Stuff

If size is not the sole determinant of value, then what is it that informs our judgements of data value in an AI context?

One factor of undoubted importance is whether or not the data is relevant to the target domain. That is to say, does the training data reflect the sort of domain in which an AI system is expected to operate? This is important. A collection of cat-related images may be of the utmost value relative to its capacity to support the development of a cat-detecting image processing system. But the very same body of training data may be of no value whatsoever if our goal is to deliver some other capability. If our goal is to build the brains behind a self-driving car, or implement a system to diagnose human diseases, then our access to a vast repository of cat-related images is unlikely to count for much, regardless of the size of the dataset and the quality of the constituent images.

The upshot is that estimates of value are inextricably linked to issues of competence. What sort of AI system do we want to build? What sort of competence do we expect the AI system to have? Our answers to these questions will influence the nature of the dataset that is required to support the acquisition of specific abilities, and this will, in turn, constrain value-related judgements, i.e., the value we assign to data and the price (political, economic, social, or whatever) that we are prepared to pay for it.

In this respect, it is interesting to note the functional specificity of many contemporary AI systems, especially deep learning systems. Despite a general appreciation of the importance of what is sometimes called Artificial General Intelligence (AGI), many of today's AI systems are designed to operate within rather restricted domains of interest. We thus have an emerging ecology of systems that are designed to do different things (e.g., recognize human faces, translate human speech, diagnose human illnesses, and so on); however, we currently lack a system that is able to transfer its expertise across domains—to take what it has learned in one domain and apply it in another. Ultimately, such forms of flexibility may depend on more than a commitment to data sharing and global data liberalization; they are likely to require a degree of cooperation between the developers of AI systems—a form of sharing that centres on the products of machine learning and not just the things (e.g., the data) that make such learning possible.

3.8.3 Synthetic Data

Not all the data that is used for training AI systems needs to be of the 'naturally-occurring' type. That is to say, synthetic (i.e., artificial) forms of intelligence can sometimes be brought into existence via the use of synthetic data. Consider, for example, the attempt to develop a variety of self-driving vehicles (e.g., cars, drones, submarines, and so on). One approach to the development of such vehicles is to acquire large amounts of real-world data and then train AI systems with respect to this data. We might, for example, equip human-driven vehicles with a bank of sensors and then collect data from the vehicles as they move around the road network. An alternative to this 'surveillance-oriented strategy' is to resort to systems that are able to replicate the dynamics of real-world sensory environments. This may sound like an impossible demand, but it is worth noting that computers are already able to generate photorealistic representations of a variety of visual environments via the use of game engine technology.

There are, of course, limitations to this sort of approach. In respect of game engines, for example, the approach assumes that we are able to model the target domain with sufficient fidelity to support the relevant learning process. This may not be possible in all cases, since we may not know enough about the causal structure of the relevant part of the environment that is the focus of our modelling efforts. If, for example, we want to train AI systems to 'understand' Chinese texts, and we know nothing about Chinese, then we will be hard-pressed (to say the least) to create a system that generates examples of Chinese text. And, if we were able to do this,

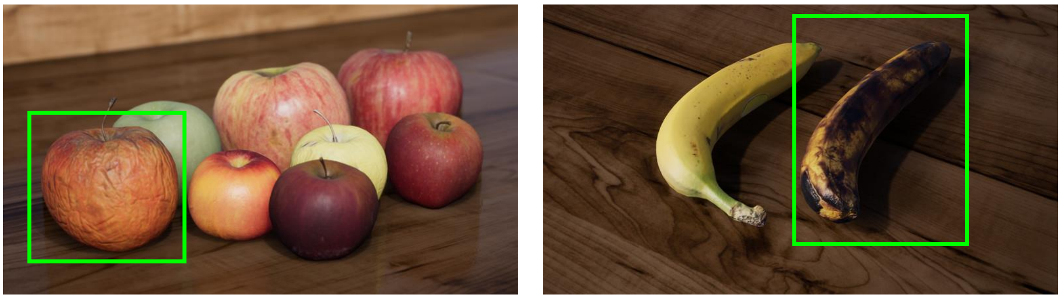


Figure 3.7: Computer-generated images of fruit with automatic annotation of target objects.

then it is not entirely clear why we would need to resort to computer-generated text as a means of training the AI system in the first place. After all, if we are able to build a computational system that understands enough about Chinese to generate semantically-coherent samples of Chinese text, then how hard can it be to build a computational system that operates in the reverse fashion—a system that understands Chinese texts by, in effect, working out what would need to be the case in order for it to generate such texts by itself? The key point here is that there are limitations to the use of synthetic data. Sometimes we do not understand enough about the target domain to build a simulation of that domain—at least one that could be used to generate useful quantities of training data.

Notwithstanding these limitations, there are a number of reasons to commend the use of synthetic data. Firstly, such approaches are divested of the sort of constraints that plague the acquisition of real-world data. We can, for example, generate as much data as we need, and we seldom have to worry about the legislative constraints that may otherwise limit our access to real-world data, especially personal data. An added bonus is that we can often augment the synthetic data with information that works in concert with the machine learning process. Note, for example, that in a purely virtual world we already know about the location of specific objects. We can therefore annotate or label our body of synthetic data with 'ground truth' information. If we want a machine vision to detect objects of a particular type, we can annotate our corpus of machine-generated images with labels that indicate the presence (and perhaps location) of the target image. Crucially, this annotation process can be done in an automatic fashion, without the need for human input. (The images in Figure 3.7, for example, depict the presence of overripe fruit in a computer-generated scene.) This addresses one of the major bottlenecks associated with machine learning, especially when it comes to the use of supervised and semi-supervised methods. The upshot is that, for at least some domains, the need for large bodies of real-world data is rendered unnecessary as a result of our capacity to build simulations of the target domain.

A nice demonstration of this approach stems from work by Johnson-Roberson et al. (2017). They show how data obtained from a purely virtual world—namely, the world depicted in the game, *Grand Theft Auto*—can be used to train a deep learning network to detect the presence of car-like objects within a visual scene. Interestingly, they report that a deep learning system trained on synthetic data outperforms a system trained on (human-annotated) data obtained from the real-world. This highlights the potential value of synthetic data when it comes to machine learning applications. It also provides the basis for empirical studies that seek to advance our understanding of what it is that makes a given body of data suitable for machine learning, and thus valuable from an AI perspective.

Synthetic data has proven to be increasingly important in AI research. For example, the US self-driving car company Waymo announced that, in a single year, cars had driven 2.5 billion miles in virtual simulators compared with only 3 million miles of real-world roads (see Madrigal, 2017). In 2017, one of the company's executives commented that "The vast majority of work done—new feature work—is motivated by stuff seen in simulation" (cited in Madrigal, 2017).

3.8.4 'Zero' Data

When it comes to issues of data-driven intelligence, it is worth reminding ourselves that not all forms of AI require access to pre-existing bodies of data. Sometimes the data required for the acquisition of intelligence can be generated as part and parcel of the learning process. A nice example of this is provided by AlphaGo. AlphaGo, recall, is a system developed by Google Deepmind to demonstrate the power and potential of recent advances in machine learning, specifically deep learning. The first version of the system, developed in 2015, was trained on real-world data derived from human Go matches. In 2017, however, Google produced a second version of the system, dubbed AlphaGo Zero. The interesting feature of this system is its learning regime. Rather than rely on human-generated data, AlphaGo Zero starts by knowing the rules of the game (Go, in this case) and then learns by playing against itself (Silver et al., 2017). There is, of course, a degree of data dependence here, since the system still requires some body of data to inform the learning process. Crucially, however, the training data is actively generated by the system itself as part of its attempt to learn about the target domain. This might be thought an unlikely recipe for success; nevertheless, the Zero (data) version of the AlphaGo system has proven to be remarkably successful. When matched against one of its predecessors, namely, AlphaGo Lee—a system that had previously beaten the 18-time world champion Go player, Lee Seedol—AlphaGo Zero was able to achieve a stunning victory, beating its predecessor by 100 games to zero (Silver et al., 2017).

Such results highlight the power and potential of what we might call a 'zero data' approach to machine learning, an approach that dispenses with the need to acquire substantive amounts of data in advance of the actual learning process. Clearly, this approach will not be suitable for all forms of machine learning—it is not intended as a blanket solution to the problems confronting the effort to build state-of-the-art AI systems. Nevertheless, the success of AlphaGo Zero is an important reminder of the fact that not all forms of synthetic intelligence (including, in this case, ones that succeed in grabbing global media headlines) rely on access to large bodies of high-quality, human-generated data. This should give us pause when it comes to a consideration of the strategic importance of a nation's data assets to the development of world-leading AI solutions. If some forms of globally-significant synthetic intelligence can be developed in the absence of human-generated data, then why assume that a nation's AI potential is always undermined as a result of (e.g.) data protection policies that limit its access to such data?

4 — The Social Credit System

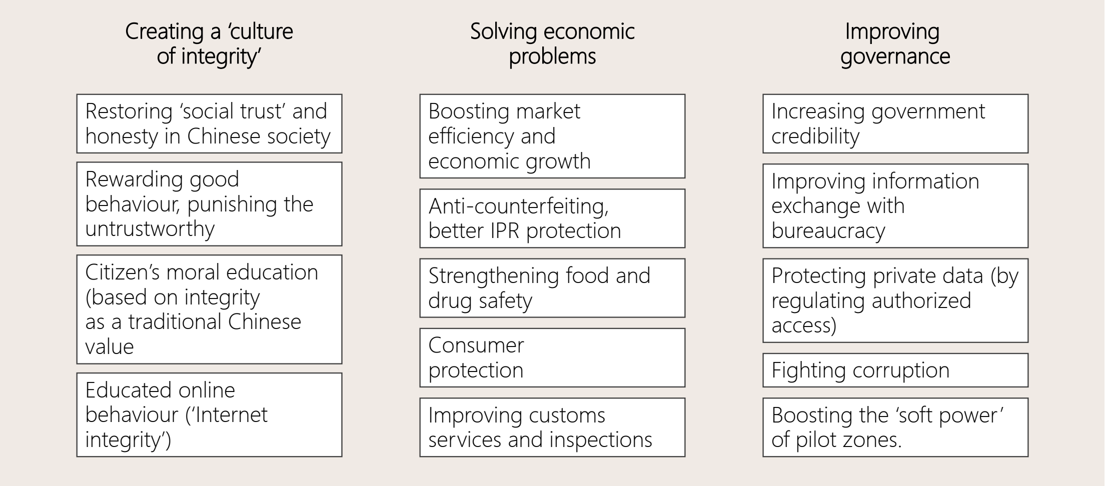
Social integrity is the foundation of the building of the social credit system. Only if there is mutual honest treatment between members of society, and only if integrity is fundamental, will it be possible to create harmonious and benign interpersonal relationships and possible to promote the progress of society and civilization, and realize social harmony, stability and long-term peace and stability.

—State Council of the People's Republic of China (2014)

No government has a more ambitious and far-reaching plan to harness the power of data to change the way it governs than the Chinese government. Its SCS, laid out in a plan released in 2014 (State Council, 2014)—the Social Credit Plan (SCP)—aims to extend financial credit scoring systems (such as those used by financial institutions in the US) to the management of social behaviour. The plan is presented as a means to make government decisions (e.g., in relation to state funding) transparent, but it is also clear that the Chinese leadership views the SCS as a tool of social governance. Under the proposed vision, a variety of social actors (e.g., individual citizens) will be assigned a social credit score, calculated from records of past behaviour. Such a score is then poised to function in the manner of a reputation metric—informing decisions made by other social actors (e.g., government agencies) about the distribution of social rewards and punishments.

Setting aside the various ethical debates that surround the SCS, it is clear that the plan confronts a number of practical challenges. In order to function as an effective constraint on behaviour, for example, a social credit score must be easily accessible and widely available. It should, in essence, stand ready to influence thought and action in a broad array of social contexts. It is no good calculating a social credit score if such a score cannot be used to drive processes of social evaluation. Another issue relates to trust and reliability. In general, a social credit score must serve as a reliable source of information about an individual. If citizens feel that the SCS is prone to error, oversight, and malign intervention then citizens may simply choose to ignore social credit information—a social credit score, in this case, is apt to be regarded as an unreliable (and thus uninformative) guide to an individual's past behaviour and their future conduct.

As evidenced by these issues, personal data lies at the heart of the SCS. The aim of the SCS is to exploit advances in data collection, data storage, and data processing capabilities as a means of shaping individual behaviour and thereby altering the dynamics of social variables (e.g., measures of social harmony or stability). In short, the SCS embodies a data-driven approach to social engineering. It seeks to use a combination of advanced technology and big data processing



Creating a 'culture of integrity'

Restoring 'social trust' and honesty in Chinese society

Rewarding good behaviour, punishing the untrustworthy

Citizen's moral education (based on integrity as a traditional Chinese value

Educated online behaviour ('Internet integrity')

Solving economic problems

Boosting market efficiency and economic growth

Anti-counterfeiting, better IPR protection

Strengthening food and drug safety

Consumer protection

Improving customs services and inspections

Improving governance

Increasing government credibility

Improving information exchange with bureaucracy

Protecting private data (by regulating authorized access)

Fighting corruption

Boosting the 'soft power' of pilot zones.

Figure 4.1: Goals of the SCS according to Chinese media reports (source: Ohlberg, Ahmed, & Lang, 2017).

to alter the properties of a social system (in this case, Chinese society). No doubt part of the motivation for the SCS is political, in the sense that it reflects a particular view as to how society should be governed. It also, of course, works in the interests of the Chinese leadership, helping to ensure the continued survival of the CCP. Finally, the SCS is broadly compatible with China's technological ambitions. In particular, to achieve its goals, the SCS requires the widespread adoption of data gathering technologies, as well as the ability to process vast amounts of personal data. All this speaks to the notion of data advantage that was discussed in Section 3.4. In effect, the SCS provides incentives for data collection and improvements data processing capability. Both of these work in concert with China's existing data-oriented strengths to help it achieve its AI ambitions.

4.1 Goals and Motivation

At a general level, the SCS is intended to tackle problems caused by a lack of honesty or sincerity in commercial, government, social, and judicial affairs. The SCS seeks to address these problems by implementing a system of rewards and punishments based around the use of personal data. Some of the goals of the SCS, as mentioned in Chinese media reports, are presented in Figure 4.1.

While the State Council's 2014 *Planning Outline for the Construction of a Social Credit System (2014–2020)* is typically taken as the starting point for discussions of the SCS, it is not the first time that official documents have referred to the notion of social credit. In 2007, for example, a State Council notice defined social credit as an "important institutional component of a market economy" focused on "credit records, tax payments, and contract performance" (State Council, 2007). The term "social credit," as it appears in this earlier document, does not have the same meaning as it does in the 2014 plan. What has changed is the idea that processes previously restricted to the purely economic realm can be used to influence more general patterns of social behaviour. Arguably, the rise of data-rich Internet companies was one of the inspirations for this move. As the capacity for data collection increased with the rise of social media, so did the government's vision of how that data could be used. The extent to which technological advances ought to be seen as the inspiration for the SCS is, however, debatable. For the SCS embodies a set of beliefs and principles that have long been a feature of Chinese society. These include the idea that the State is not merely in charge of legal authority, but that it should also play a role in fostering social morality (Creemers, 2018a). It also includes a commitment to the idea that

Table 4.1: The SCS supports four aspects of China's approach to autonomic governance (source: Hoffman, 2017).

Governance Objective	Definition	The CCP's Objectives
Self-Configuring	A process of readjusting automatically to respond to changing circumstances, or to support the process of 'self-healing', 'self-optimization' or 'self-protection'.	Adjustments in government structure aimed at supporting the social management process through vertical and horizontal integration, and also at controlling the power of the individuals and agencies controlling vast state security resources.
Self-Healing	A system's ability (in the reactive sense) to self-fix faults and (in the proactive sense) to predict and prevent faults.	A pre-emptive management of threat at the source. The attempt to automate this process is found through application of advanced surveillance technology to grid management.
Self-Optimization	A system's ability to dynamically optimize its own operation.	Pre-empting challenges to the Party's control via creation of a system, in which every member of society has the responsibility to participate in their own management. The attempted automation of this process is found in the SCS.
Self-Predicting	A system's ability to protect itself through perception of potential threats and prediction of outcomes of situations in the environment, and through self-configuring to minimize potential harm.	Preparation for response to emergencies of all types through more effective coordination and pre-emption. The attempted automation is found through efforts to integrate technology applications to support emergency pre-emption and response.

societies can be 'engineered' via an appeal to cybernetic principles (see Hoffman, 2017).

The latter of these two ideas (a commitment to cybernetic principles) is reflected in the way in which the SCS establishes a system of 'self-control' or 'self-correcting' feedback to social actors. The core idea is that the SCS provides a means by which the mechanisms for social control are embedded within the fabric of society, thereby reducing the need for more direct forms of government intervention. Such an approach, which Hoffman (2017) dubs an autonomic approach to governance (see Table 4.1), is likely to be a highly efficient form of social governance, a property that is no doubt important given the scale and complexity of Chinese society.

Aside from issues of efficiency, an autonomic approach to social governance comes with an added bonus: it reduces the need for governments to resort to coercive methods as a means of controlling behaviour. Thus, rather than seek to control individual behaviour via the use of strict legislation and robust policing, the SCS delegates control to the social environment. This is important, because it detracts attention from the role of the government in constraining or limiting individual action. From a psychological perspective, this no doubt influences the way that individuals interpret their behaviour and thus (perhaps) change their attitudes (see, for example, Cooper, 2007). More importantly, however, the SCS avoids the need for the sort of coercive tactics that might otherwise attract the opprobrium of the international community. This point is nicely made by Creemers (2018b):

> The problem of coercion damaging the Party's image and causing a domestic and international backlash is overcome by modifying the space for action, and subtly employing less prominent means, including technology and code, either to nudge individuals towards compliance or to make it impossible for them to engage in

> undesired activities. These powers are primarily seen, and utilized, as a tool to prevent organized opposition and dissent. (Creemers, 2018b, p. 269)

Inasmuch as the SCS embodies values that long have been a feature of Chinese society, then we might see the Internet as merely supporting the expression of pre-existing, albeit latent, approaches to social governance. There is, in this sense, nothing 'new' about the SCS—it does not, for example, represent a shift in China's view of what society is or how it should be controlled. But the Internet has not merely served as an enabler for the SCS, it has also highlighted the need for greater government control over individual and commercial conduct. With the advent of the Internet, reports of fraud, commercial scandals, and individual misconduct could be easily communicated at a rate that was difficult for the Party's propaganda machine to control. The Internet also provided new opportunities for the public to express dissatisfaction with government bodies. Inadequate responses to emergency situations were easily documented, as were incidences of government corruption (see Creemers, 2017, for examples). Given that such reports threatened to erode the trust that people placed in government institutions, it was important for the Chinese leadership to act.

Perhaps, then, we can understand the SCS as part of a broader effort to resolve some of the challenges thrown up by the Internet. In this sense, the Internet is just as much a catalyst of government policy as it is an enabler of government policy. Such a vision suggests a degree of interdependence between China's approach to cyber, social, and data governance.

4.2 Social Credit Systems

Although the SCS is commonly portrayed as a single, unified system that operates across the entire expanse of Chinese society, there is, at present, no *single* SCS in China. Instead, there are a variety of systems, many of which operate in specific geographical regions. For present purposes, we can distinguish three categories of SCS. These are national platforms, local pilot projects, and commercial credit systems. Subsequent sections provide a brief overview of these systems.

4.2.1 National Platforms

National platforms are systems that operate at the national level, with data drawn from the entire expanse of mainland China. This does need not mean that the systems themselves are distributed. The data associated with a national-level system may be stored in a particular location, and thus centralized; however, the 'data catchment area' for such systems is typically one that is coterminous with the borders of mainland China.

According to Liang et al. (2018), China is building at least five national data platforms to store and evaluate credit-related data. These are the National Credit Information Sharing Platform (NCISP), Credit China, the Credit Reference Center (CRC), the National Enterprise Credit Information Publicity System (NECIPS), and the List of Dishonest Persons Subject to Enforcement (see Table 4.2).

Clearly, not all of these national-level platforms have the sort of scope typically associated with the SCS. What makes the SCS the target of Western interest (and criticism) is the nature of the data that feeds into social credit scores. In particular, the ethical sensibilities of Western commentators are typically offended by the potential scope of social credit scores—the idea that social credit scores extend to (e.g.) the expression of political sentiments in social media posts and thereby undermine individual autonomy and freedom of speech. Relative to these concerns, it should be clear that most of the platforms listed in Table 4.2 have a much more limited scope. The CRC, for example, functions in a manner similar to conventional (finance-related)

Table 4.2: National social credit platforms (source: Liang, Das, Kostyuk, & Hussain, 2018).

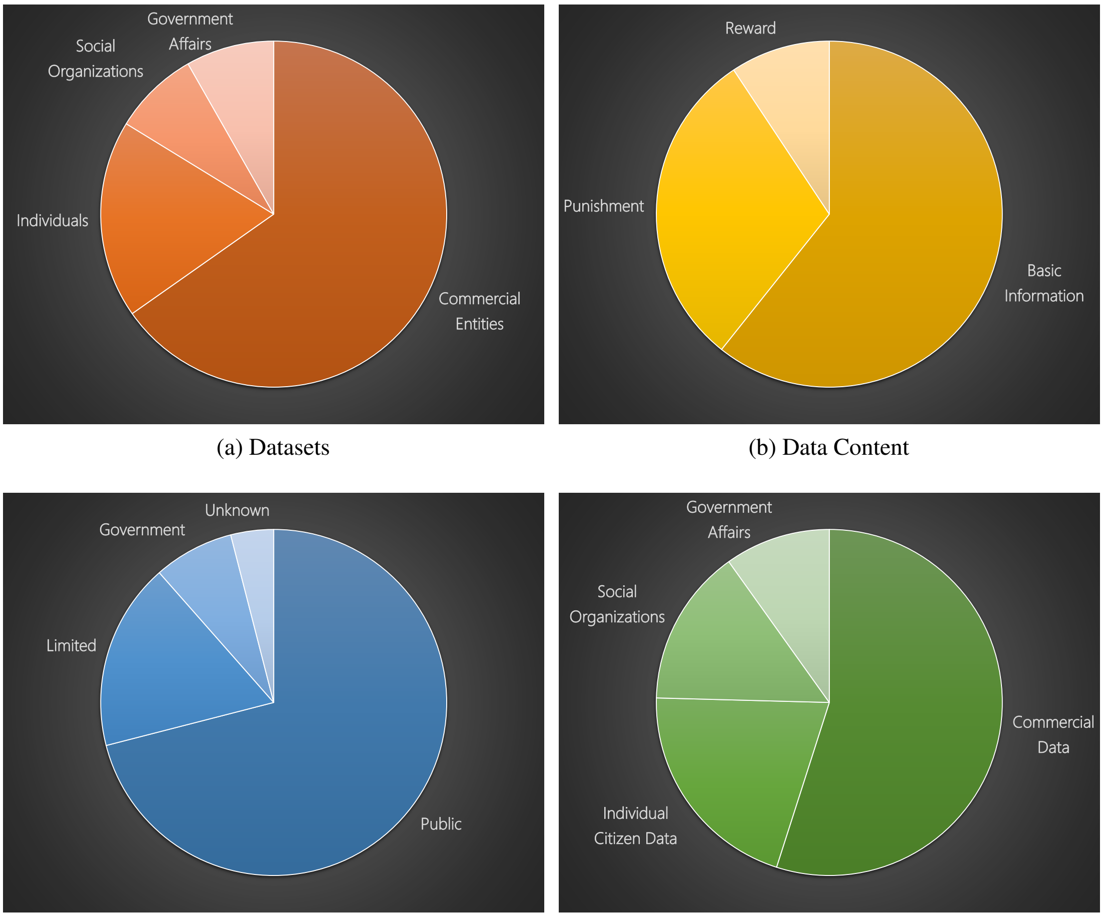
Platform	Date	Stakeholders	Focus
Credit Reference Center	2006	People's Bank of China	Public; Commerce
The Blacklist of Trust-breaking Platform	2013	The Supreme People's Court	Public; Commerce; Government
National Enterprise Credit Information Publicity System	2014	State Administration for Industry and Commerce	Commerce
Credit China	2015	National Development and Reform Commission; People's Bank of China; State Information Center; Baidu	Public; Commerce
National Credit Information Sharing Platform	2015	National Development and Reform Commission	Public; Commerce; Government

credit agencies. According to Shen (2019), the CRC provides individual credit reports that contain information about personal loans, mortgages, credit card use, delayed payment records, civil judgment records, unpaid utility fees, administrative penalties, and so on. As noted by Shen (2019), such reports do not provide a single credit score (or credit rating) for individuals and enterprises, and they thus differ from the credit reports delivered by credit agencies in Western societies.

A platform that is more closely aligned with the vision of the SCS is the NCISP. This platform has been glossed as "the data backbone of the Social Credit System" (Meissner, 2017, p. 6). The NCISP is currently undergoing development by the National Development and Reform Commission (NDRC). It is connected with 42 central government agencies, 32 local governments, and 50 market actors (see Liang et al., 2018).

The datasets collated by the NCISP involve both public data (e.g., administrative licenses) and private data (e.g., citizen names and ID numbers). In total, the NCISP houses 400 datasets (Liang et al., 2018). Among these 400 datasets, about two-thirds ($N = 261$) focus on firms and commerce, while around a fifth ($N = 74$) collect data from individuals. The remaining datasets cover social organizations ($N = 32$) and government affairs ($N = 33$) (see Figure 4.2a). Around two-thirds of the datasets ($N = 244$) gather basic information, including the names and addresses of firms/persons, social credit numbers, and ID numbers. This data is mainly used to identify social actors. A third of the datasets ($N = 119$) contain "trust-breaking information," such as information about prior punishments, bad credit records, and crime records. Finally, 37 datasets include reward-related information, such as information about volunteer services, qualifications, and awards (see Figure 4.2b).

In terms of data content, the NCISP uses at least 537 variables to measure commercial firms, individual citizens, social organizations (e.g., non-governmental organizations), and government agencies. Specifically, a little over half (295) of the variables are used to collect information about commercial entities. Some variables focus on basic information like the firm's name, location, legal representatives, website, social credit numbers, phone numbers, shareholders, and product information. Other variables emphasize rewards and honours, punishments, and criminal records. Around a fifth of the variables ($N = 110$) relate to individual citizens. This includes information about names, ID numbers, trust-breaking behaviours, certificates, and administrative penalties. Note that this does not mean that the NCISP collects 110 variables from everyone in China. Instead, many variables are obtained from specific groups of people, such as lawyers, teachers, and students. In addition, 79 variables are employed to gather data from social organizations, including information about addresses, service sector, administration



(a) Datasets

(b) Data Content

(c) Data Accessibility

(d) Data Variables

Figure 4.2: Overview of the NCISP (source: Liang, Das, Kostyuk, & Hussain, 2018).

registration, and trust-breaking activities. Finally, government affairs make use of 53 variables. These focus on administrative licenses and penalties (see Figure 4.2d).

Among the 400 datasets, 384 are tagged as belonging to one of three sharing categories: public sharing, limited sharing, and inter-government sharing. Specifically, three-quarters of the datasets ($N = 284$) are publicly available, meaning that the public can access these datasets. By contrast, around a fifth of the datasets ($N = 70$) (mostly those containing punishment-related information) are limited for sharing; however, it is unclear who has access to these datasets. Finally, the remaining 30 datasets are only shared among government agencies. These datasets include those pertaining to things such as student credit and administrative penalties (see Figure 4.2d).

Of the 42 central government agencies involved, NDRC provides 73 datasets (68 commercial datasets and five citizen-focused datasets), which makes it the most prominent contributor. The Ministry of Industry and Information Technology (MIIT) is another critical agency, sharing 34 datasets, followed by the Ministry of Agriculture (28 datasets), the National Health and Family Planning Commission (27 datasets), the Ministry of Transport (24 datasets), and the Ministry of Housing and Urban–Rural Development (24 datasets). Overall, these six agencies supply a little over half ($N = 210$) of the NCISP's datasets. In contrast, the Ministry of Culture (three datasets), Cyberspace Administration of China (CAC) (two), Securities Regulatory Commission (two), and the State Administration for Industry & Commerce (one) only provide limited data to the

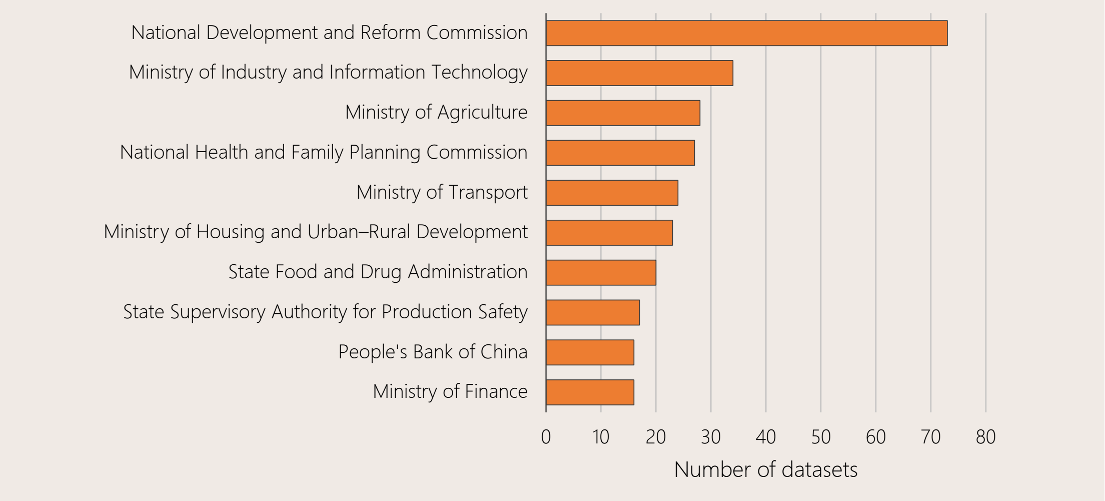


Figure 4.3: Top 10 data providers for China’s NCISP (source: Meissner, 2017).

system (see Figure 4.3).

The national-level platforms described in Table 4.2 should not be viewed as independent, stand-alone systems. Rather, such systems participate in complex webs of information exchange, and it is these larger webs that are likely to support the functionality of the SCS. Some insight into the nature of the information flow patterns between national platforms is provided by Liang et al. (2018):

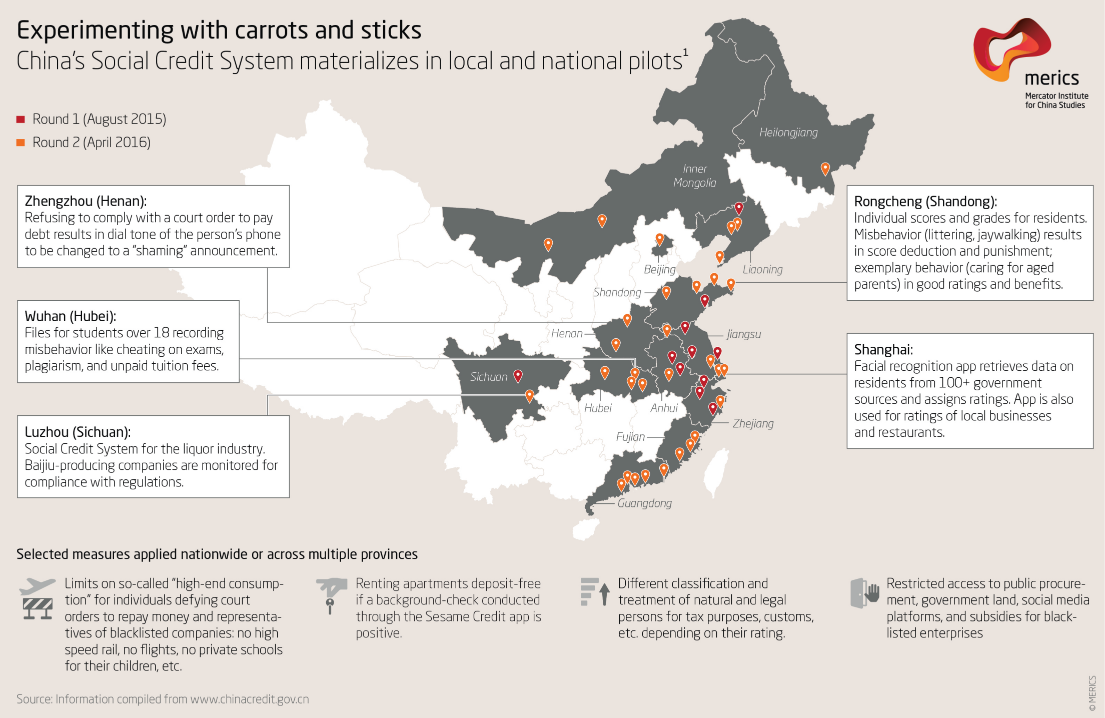
> If a subject displays serious trust-breaking activities, then NCISP will directly share relevant data with Credit China—another national credit data platform developed by NDRC and the People’s Bank of China (PBoC). Otherwise, this system will establish a Focus Group List including those who are assumed to be trust-breaking subjects but who have yet to meet the standards of the blacklist. If this subject is mentioned by three different sources, then NCISP will move this subject into a Big Data Warning List for further investigation. In the final phase, NCISP will share data with Credit China, and the later [*sic*] will announce the red list and blacklist for the public. Meanwhile, local credit websites and sectoral websites will also release rewards and punishments information. Evidently, these data platforms play different roles in this process: while NCISP emphasizes information gathering and evaluation, Credit China focuses on publicizing the blacklist. (Liang et al., 2018, p. 433)

4.2.2 Local Pilot Projects

Although China lacks an overarching SCS at the time of writing, the Chinese government has authorized a number of SCS projects to be implemented at the municipal and provincial level (see Figure 4.4). These local pilot projects no doubt serve as an important source of information about the challenges confronting the roll-out of a national-level SCS.

Unfortunately, much of the evidence in support of particular SCSs is anecdotal and there has been little attempt to quantify their social effects. At the very least, there is little evidence to suggest that such analyses (if they have been undertaken) are available to a Western audience.[6]

[6] This is not to say that such results are unavailable to the China’s scientific community and/or government. The results of such studies may not have been disclosed, or they may have been reported in Chinese academic journals. This is an area for future research (see Section A.4.1).


Experimenting with carrots and sticks

China's Social Credit System materializes in local and national pilots[1]

- Round 1 (August 2015)
- Round 2 (April 2016)

Zhengzhou (Henan): Refusing to comply with a court order to pay debt results in dial tone of the person's phone to be changed to a "shaming" announcement.

Wuhan (Hubei): Files for students over 18 recording misbehavior like cheating on exams, plagiarism, and unpaid tuition fees.

Luzhou (Sichuan): Social Credit System for the liquor industry. Baijiu-producing companies are monitored for compliance with regulations.

Rongcheng (Shandong): Individual scores and grades for residents. Misbehavior (littering, jaywalking) results in score deduction and punishment; exemplary behavior (caring for aged parents) in good ratings and benefits.

Shanghai: Facial recognition app retrieves data on residents from 100+ government sources and assigns ratings. App is also used for ratings of local businesses and restaurants.

Selected measures applied nationwide or across multiple provinces

- Limits on so-called "high-end consumption" for individuals defying court orders to repay money and representatives of blacklisted companies: no high speed rail, no flights, no private schools for their children, etc.
- Renting apartments deposit-free if a background-check conducted through the Sesame Credit app is positive.
- Different classification and treatment of natural and legal persons for tax purposes, customs, etc. depending on their rating.
- Restricted access to public procurement, government land, social media platforms, and subsidies for blacklisted enterprises

Source: Information compiled from www.chinacredit.gov.cn

Figure 4.4: SCS pilot projects at the level of local government (source: Ohlberg, Ahmed, & Lang, 2017).

One system that has received plaudits is a SCS operating in Qingzhen, a city in Guizhou province. According to Dai (2018) the system uses 1,000 indicators to assign a point value to its citizens. It also includes peer evaluations and community monitoring, drawing uncomfortable parallels with the "Nosedive" episode of the television show *Black Mirror*. Dai notes a number of positive changes brought about as a result of the introduction of the Qingzhen SCS. These include increases in Gross Domestic Product (GDP), residential income, flow of investment, an improvement in social order and "harmony." Encouraged with these results, Qingzhen officials announced in 2017 that they planned to take the SCS to the next level with a major technological upgrade featuring blockchain technology (Dai, 2018).

Many of the local pilot projects that have been discussed in the English literature are those that operate at the level of Chinese cities. At the end of 2017, the NDRC and the PBoC designated twelve "model cities" from which important lessons could be learnt (National Development and Reform Commission, 2017). These cities—Hangzhou, Nanjing, Xiamen, Chengdu, Suzhou, Suqian, Huizhou, Wenzhou, Weihai, Weifang, Yiwu, and Rongcheng—were acknowledged for their achievements in a variety of areas. According to Creemers (2018a):

> Ximen [*sic*], a prosperous port city on the Taiwan Straits, was listed for its achievements in expanding the credit structure to the processing of imported and exported goods, as well as the efforts it had made to improve healthcare, education and legal services through social credit. Suqian, a smaller city in the province of Jiangsu that has attracted some of Shanghai's ICT industry, gained plaudits for the quality of its technical infrastructure and security arrangements, as well as its locally developed "Credit Plus" model, which actively stimulates local businesses and organizations to make public credit commitments... Weihai, a seaport in Shandong, was commended for the efficient way in which individuals could access the credit information system,

> and its initiative in producing clear lists of reward and punishment measures. It also had introduced a ranking system for businesses, dividing them into four tiers based on their compliance with market regulations. Another city in Shandong, Rongcheng, entered the list on the strength of its tiered system for individuals, ranking them in six classes, from AAA to D. (Creemers, 2018a, pp. 18–19)

Among the more widely discussed pilot projects is the Rongcheng SCS. One of the notable features of this system is its high social coverage. All permanent residents, non-permanent residents, self-employed individuals, enterprises, social organisations, and villages are included in its database (C. F. Shen, 2019).

The system is also notable for its rating scheme. As mentioned in the above quotation from Creemers (2018a), the Rongcheng system ranks social actors according to six categories. All actors are initially assigned a rating of 1000 points, which translates to an A category rating. From there, ratings can rise to AAA or fall to B, C, or D. The way in which these ratings are affected by social data is not terribly clear, although Rongcheng has created a catalogue of activities—the *Catalogue of Soliciting Social Credit Information*—which purportedly details over 600 economic and social activities that may affect social credit ratings (Ojeda, 2019; C. F. Shen, 2019).

A third feature of the Rongcheng system is its emphasis on rewards. In this respect, the system is deemed to be more successful than SCSs that focus exclusively on punishments, or mixed-model SCSs that seek to combine rewards and punishments. As noted by Síthigh and Siems (2019), this may point towards something of a general principle in the design of SCSs: Systems that emphasize reward are likely to be more effective (or, at any rate, more acceptable) than those that emphasize punishment or that feature a common emphasis on reward and punishment. This principle may be something that is recognized by the central government. The 2014 SCP, for example, speaks of "Perfect operational mechanisms for the social credit system with rewards as the focus point."

Interestingly, not all of the 'rewards' meted out by SCSs are ones that can be measured in monetary terms. The Rongcheng system, for example, relies on the use of public displays to recognize those who have distinguished themselves within the system (see Figure 4.5).

Some of the features of the Rongcheng system are shared by a SCS that operates in Shanghai. As with the Rongcheng SCS, Shanghai's SCS focuses on rewards. That is to say, the system only provides rewards for good credit ratings; there are no explicit punishments for bad ratings (although such punishments may emerge as a result of changes in one's social standing). The Shanghai SCS features the use of a smartphone app, called "Honest Shanghai." Residents can opt to voluntarily download and sign up for this app by providing their identification number from their resident identity card or by using facial recognition technology (Creemers, 2018a). The application draws upon from 3000 pieces of information from nearly 100 public authorities (Schmitz, 2017). It relies mainly on facial recognition software to identify and locate chunks of personal data associated with an individual across multiple government platforms. This data is then collected and integrated to produce a customised credit report on individuals and businesses. One of the incentives for participating in the system is the capacity to check the credit report of local businesses, such as restaurants and international trading businesses. The app also allows registered users to obtain a credit rating for themselves, which yields a result of "very good," "good," or "bad." According to Síthigh and Siems (2019), this rating is based on data the Shanghai government has collected about each citizen; details of the algorithm are, however, not available.

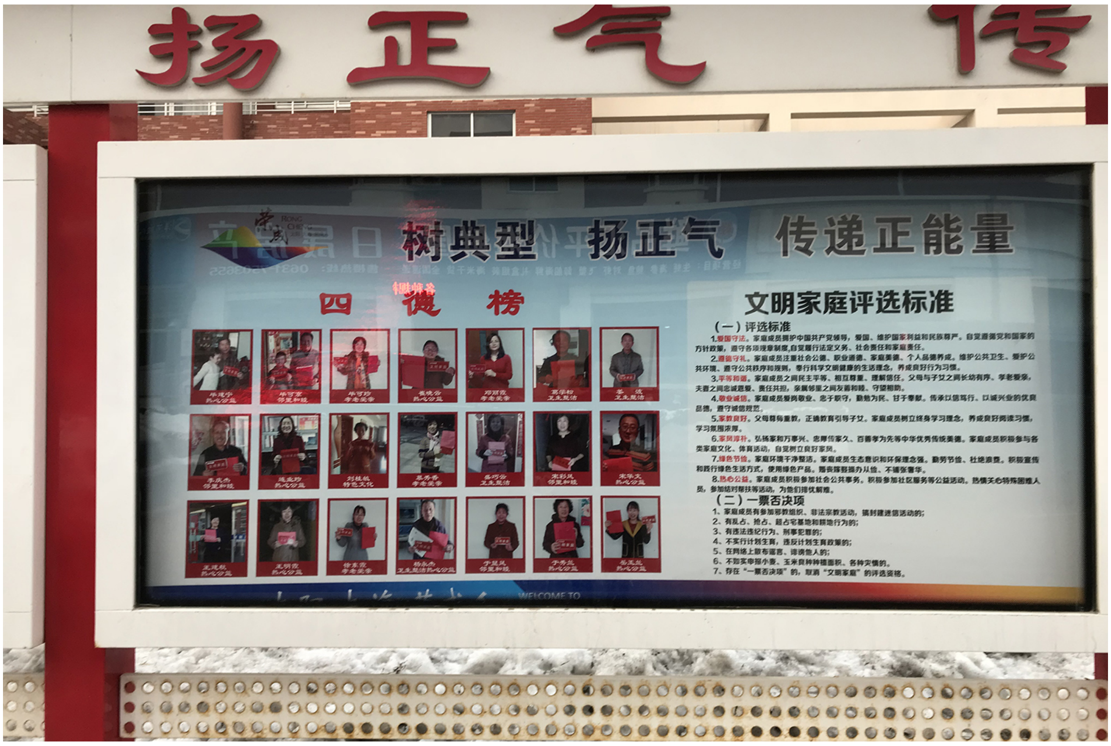


Figure 4.5: Public displays of creditworthy individuals in Rongcheng (source: Mistreanu, 2018).

4.2.3 Commercial Systems

In addition to government-led SCSs, China also features SCSs run by commercial companies. Many of these systems have their origins in e-commerce systems. As noted by Dai (2018):

> Drawing on the massive consumer transactional data accumulated through their business operations, e-commerce firms started to conduct credit reporting and ratings on their users, which allowed them to optimize operations, regulate platform activities, facilitate credit-based transactions, and eventually expand into core financial services. (Dai, 2018, p. 16)

Of all the commercial systems that currently operate in China, Sesame Credit, developed by Ant Financial Group, an affiliate of the e-commerce giant Alibaba, is the system that has received the most attention. Sesame Credit calculates a score between 350 points and 950 points based on the following data categories (the percentages in parentheses indicate the weighting that is assigned to each category) (Creemers, 2018a; C. F. Shen, 2019):

- credit history, or records of past credit repayments (35%);
- behavioural trends, referring to someone's conduct when making purchases, processing payments, settling accounts and managing their finances (25%);
- fulfilment capacity, referring to personal assets and income (20%);
- personal information, referring to the amount of verifiable and reliable information about registered members (15%); and
- social relationships, referring to the extent to which one interacts with good friends and behaves in a friendly manner on the platform (5%).

The nature of these data categories highlights one of the reasons why Sesame Credit ought to be regarded as a SCS (and perhaps why it is sometimes confused with the SCS proposed by the Chinese government). Thus, in addition to incorporating information about an individual's financial capacity and credit history, the system also seeks to incorporate information about an individual's (online) social conduct and (online) social relationships. To a Western observer, the inclusion of such information is apt to lead to worries about social manipulation and control. It is, however, a mistake to view Sesame Credit in a historical and cultural vacuum. In particular, it is important to bear in mind some of the challenges faced by China's e-commerce systems as they emerged in the early 21st Century. Such challenges included the absence of the sort of payment mechanisms and financial-related reporting structures that were already well-established in many Western countries. In all likelihood, it was the inaccessibility of this particular category of personal data (i.e., financial-related data) that motivated companies to broaden the scope of their credit scoring systems. As noted by Creemers (2018a):

> ...Sesame solved several problems that Alibaba faced in the development of its e-commerce business. The most important one of these concerned payment and creditworthiness. China was, at that time, still a largely cash-based economy, where few individuals held credit cards. Alibaba's model as an intermediary for purchasers and sellers required it to be able to facilitate payment. Without external broad-based providers, Alibaba decided to set up a mobile payment system of its own, which became known as Alipay. This, in turn, required Alibaba to be able to assess customers' financial creditworthiness. In the absence of a recorded financial history, they turned towards other potentially useful proxies. (Creemers, 2018a, p. 23)

The primary purpose of Sesame Credit is to assess an individual's creditworthiness, i.e., their suitability for financial loans and other forms of financial credit. This makes sense given that Sesame Credit is run by a financial institution (i.e., Ant Financial). As Sesame Credit has grown in popularity, however, its influence has expanded to applications beyond the purely financial domain. Some of the benefits associated with high Sesame Credit Scores include the following (Creemers, 2018a; C. F. Shen, 2019; Síthigh & Siems, 2019):

- expedited procedures when applying for Singaporean and Schengen visas;
- reductions or waivers for deposits on a host of products and services, including mobile phones, hotels, and bicycle rentals;
- access to fast lanes at airport security; and
- deposit waivers for access to local government services in areas such as healthcare, social housing, and public services.

There are also reports of credit scores being used to inform romantic decision-making. According to Hatton (2015), for example, China's biggest matchmaking service, Baihe, has teamed up with Sesame Credit to promote users with good credit scores, giving them prominent spots on the company's website.

Such applications highlight the value of credit scores as a form of 'reputational currency'. They also tell us something about the importance of trust-related information in a variety of decision-making contexts. Trust is, of course, an important feature of social life. In general, it helps to know how an individual will behave in specific situations, and such expectations are typically informed by a consideration of past behaviour (as well as information about personal characteristics). In this sense, it is perhaps not surprising that trust metrics developed to support decision-making in one domain (e.g., the financial domain) should be incorporated into other aspects of social life.

All this raises a question about the extent to which systems like Sesame Credit might be used as components of a larger, government-controlled SCS. Sesame Credit has typically denied any involvement with the government in respect of its plans to develop China's SCS. In a letter to the Financial Times, for example, Sesame Credit's general manager, Hu Tao, denied that Sesame Credit was contracted to develop the government's SCS (H. Tao, 2017). Hu also stated that Sesame Credit did not share data or credit scores with the government or other third parties without a user's consent.

4.2.4 Comparing Social Credit Systems

The systems described above—national platforms, local pilot projects, and commercial systems—provide us with some insight into the features of different SCSs. This is important, because it reminds us that SCSs do not come in a single flavour. There are many different kinds of SCS that can be distinguished with respect to various features. Relative to the foregoing analysis, we can now identify some of these features. They include:

- **Social Coverage:** Who is included in the system? (The Roncheng system, recall, scores highly on this dimension.)
- **Rating Scheme:** What is the nature of the rating scheme used by the system? Does the rating scheme use ordinal, interval, or ratio variables?
- **Data Coverage:** What kind of data is collected/analysed?
- **Credit Evaluation Process:** How is a credit score derived?
- **Incentives:** What are the rewards and punishments?
- **Reward Scheme:** Does the system emphasize rewards, punishments, or a combination of rewards and punishments?

How China's SCS fares with respect to these (and other) features will determine where it lies within a much larger space of design possibilities. Ultimately, however, China's SCS is only a single point within the 'universe' of SCS designs. Thus, while China's SCS may prove unworkable or undesirable in a Western context, this does not mean that the more general notion of social credit and data-driven shifts in social behaviour is entirely without merit. Consider, for example, a system to promote fitness-related behaviours by rewarding individuals, not for the number of times they themselves frequent a gym, but for the number of times people within their social network visit a gym. Such a scheme does not seek to reward gym-going behaviour as such; what it rewards are behaviours that encourage one's peers to go to the gym. To be sure, such a system requires a certain amount of intrusion into one's private life—it requires access to details about one's social network and (of course) the number of times that specific individuals attend a gym (hopefully, to engage in some fitness-related activity). At the same time, however, the system is not intended to coerce or constrain; instead, it is designed to indirectly influence social behaviour by encouraging individuals to implement (for themselves) the sorts of social scaffolding (e.g., supportive social networks) that promote the adoption of health-related behaviours. In this case, there is little reason to see the system as an instrument of government control or oppression; it is, instead, a relatively lightweight regulatory scheme that lays the foundation for citizens to erect their own forms of behaviour-changing infrastructure.

There have been a number of attempts to characterize the various sorts of SCS that currently operate in China. One such effort is described by Síthigh and Siems (2019). They attempt to distinguish SCSs according to their "degree of interventionism," which is calculated based on the responses to the following questions:

1. **Drafter:** Is the SCS initiated by a private or public authority? Private authorities (e.g., private companies) are deemed to represent a low level of interventionism, while public authorities (e.g., national governments) represent high levels of interventionism.

2. **Aim:** Does the system have a single, specific aim, or does it have a broader set of objectives? If the aim is to regulate behaviour in a specific context (e.g., good driving behaviour) then it is deemed to represent a low level of interventionism.
3. **Scoring:** Does the SCS use single or multiple scores?
4. **Application:** How flexible is the system's application? A system with low flexibility is characterized by a common system of rewards and punishments that are uniformly applied to participants of the SCS, regardless of the presence of mitigating circumstances.
5. **Algorithm:** How transparent are the algorithms used to calculate social credit scores?
6. **Enforcement:** Who is responsible for monitoring/managing the SCS?

There have also been attempts to compare China's SCSs with financial credit systems, including those that operate outside of mainland China. Chorzempa et al. (2018) provide a nice example of this sort of analysis (see Table 4.3).

4.3 Technical Infrastructure

From an infrastructure perspective, the SCS requires the resolution of a number of challenges associated with the collection, integration, and processing of data. It also requires an ability to link particular bodies of data with a particular data subject. These requirements are recognized in the SCP. The SCP document thus talks about the development of a "big data infrastructure," which consists of a "national data sharing exchange platform," "national governance data applications," and "big data information databases." At present, the government's approach to the implementation of this "big data infrastructure" remains obscure. As noted by Liang et al. (2018), the government is likely to rely on commercial solutions for some (and perhaps all) of the components of the SCS. This highlights the importance of understanding commercial approaches to the implementation of mechanisms associated with the SCS, for example, credit scoring mechanisms, data collection mechanisms, and data sharing mechanisms.

4.3.1 Subject Identification

The SCS will require a system for the digital identification of both individuals and organizations. Considerable progress has been made with respect to the creation of unique identifiers for organizations. These are called "social credit numbers" (see Liang et al., 2018). In all likelihood, individuals will be identified by the 18-digit identifiers that are currently assigned to all Chinese citizens.

The presence of an identifier is, of course, only the first step in relating specific bodies of data to particular data subjects. In this respect, China's emerging expertise in biometric identification is likely to be invaluable in identifying subjects across a range of contexts. In addition to advances in facial recognition, led by companies such as SenseTime, China is also developing capabilities for voice identification. In 2018, the Guizhou provincial government, Tsinghua University, and Beijing-based d-Ear Technologies announced a pilot project intended to create a national database of "voice-prints" and link them to national identifiers (Zaagman, 2018, p. 41).

China is also well-placed to identify data subjects in the online environment. Of crucial importance here is China's policy of real-name registration. This requires the providers of online services to link user registration information to subject identifiers. According to Article 24 (Paragraph 1) of the 2017 *Cybersecurity Law*:

> Network operators handling network access and domain name registration services for users, handling stationary or mobile phone network access, or providing users with information publication or instant messaging services, shall require users to

Table 4.3: Comparison of social and financial credit systems (source: Chorzempa, Triolo, & Sacks, 2018).

Dimension	Sesame Credit	Credit Reference Center	Social Credit System	United States Credit Bureaus
Goal	Expand consumer credit, drive users to Alibaba products.	Expand access to finance and lower lending risk.	Use creditworthiness to strengthen trust and order in government, throughout life in China.	Collect and leverage data to price loans effectively and evaluate credit.
Type	Private	Public	Public	Private
Operated by	Alibaba Affiliate Ant Financial, regulated by the People's Bank of China.	People's Bank of China.	National Development and Reform Commission as lead, other government departments for areas of jurisdiction.	Equifax/Experian TransUnion and FICO.
Who is rated/scored?	Individual Alipay users must opt-in; companies can be rated but on a different point scale.	Automatic inclusion of 900 million records on individuals but no scores; company information also collected.	No individual choice; data automatically collected; and strict user consent, processing, and sharing provisions may be imposed.	Anyone the credit bureau can get data on. No opt-out. Bureaus collect data and produce reports. Joint ventures by credit bureaus called VantageScore and FICO's Score are main scores.
Data sources	Mostly Alibaba data: shopping, payments, other sources users share.	Financial institutions regulated by the People's Bank of China; some online lenders like Ant Financial.	Government data from departments at all levels, trains, courts, etc.; some from private firms.	Public records and lenders participating in the reporting system.
Output	Single score designed to estimate the likelihood of loan default.	Credit report.	Laws, databases, systems for data exchange, punishments, and incentives; scores possible in future.	Credit report and single score designed to estimate likelihood of loan default.
Consequences of low or bad rating	More expensive credit from Ant Financial, down payment not waived for some rentals, less access to Alibaba services.	Financial institution rejects credit application, requires collateral, or charges higher loan interest.	Expanding rewards and punishments; including lost access to government subsidies, inability to purchase plane/train tickets.	Inability to borrow or open a credit card, rent an apartment, get hired for many jobs, and more.

> provide real identity information when signing agreements with users or confirming the provision of services. Where users do not provide real identity information, network operators must not provide them with relevant services. (Standing Committee of the National People's Congress, 2017)

The real-name registration requirement applies to more than just the registration of Internet domain names. In February 2015, the CAC mandated a real-name registration system for all account-based online information services (Cyberspace Administration of China, 2015). This includes accounts for social media platforms such as WeChat. Real-name registration is also required for the purchase of smartphones, which provides an additional means of tying online activities to specific individuals.

4.3.2 Data Collection

At present, it is unclear what kinds of personal data will be incorporated into China's SCS. This makes it difficult to assess the extent to which China's existing technological infrastructure is able to support the data collection process. Given the foregoing discussion of the real-name registration, it is likely that China has ample capacity to monitor patterns of activity online. There is also evidence to suggest that China is using AI technologies to support the analysis of online activities. According to research by Knockel et al. (2018), for example, one of China's most popular social media systems, WeChat, relies on sophisticated image processing technology to analyse the text embedded in image posts—WeChat users sometimes resort to such posts in an effort to circumvent the censorship schemes that are applied to conventional text.

In addition to its efforts in the online realm, China is widely believed to have an extensive surveillance capability. According to one report, China had a network of 176 million surveillance cameras in 2017, which was expected to increase to 626 million by 2020 (Grenoble, 2017). With the integration of facial recognition technology, and other AI-based capabilities, including crowd-based analyses and human action analysis, China is poised to create one of the world's most advanced social surveillance systems, enabling it to cross-reference real-world surveillance footage with other kinds of digital data.

Some insight into China's surveillance capabilities were revealed by a data leak emanating from a Shenzhen-based company, called SenseNets (L. Tao, 2019; Yang & Murgia, 2019). SenseNets specializes in facial recognition, crowd analysis, and human verification. It forms part of China's Skynet Project—a mass surveillance effort driven by over 20 million street cameras. According to media reports of the data leak, SenseNets collected nearly 6.7 million GPS coordinates in one database in a 24-hour period. This location-related data was matched to the names of individuals, along with ID numbers, photos, employer names, and textual data such as "mosque," "hotel," "internet cafe" and other places where surveillance cameras were likely to be installed (L. Tao, 2019; Yang & Murgia, 2019).

It is easy to assume that this emerging surveillance infrastructure will be used to support the objectives of the SCS. For the time being, however, there is no hard evidence to suggest that China actually plans to do this. Such systems may be incorporated in the SCS; however, they may also function as somewhat isolated systems that monitor behaviour in geographically- and socially-specific contexts. One example of this is the use of facial recognition systems to enforce compliance with road traffic conventions at pedestrian crossings (see Section 4.6). Such systems are already in use within a number of Chinese cities.

4.3.3 Data Aggregation

Data aggregation refers to the attempt to integrate and merge multiple bodies of data for the purpose of calculating a social credit score. Without such forms of aggregation, it is unclear how the SCS can be made to work. It is clear that China's government recognizes the importance of data aggregation. China's 13th Five Year Plan for National Informatization, released in 2017, says that the country should create a "national data resource system" responsible for forming connections between ministries and promoting interdepartmental information sharing (State Council, 2016a). The 2014 SCP also calls for progress in supporting the exchange of credit information between government departments. Finally, data aggregation is seen to be relevant to other activities, including the development of AI capabilities. In the 2017 *New Generation Artificial Intelligence Development Plan*, for example, we see an explicit recognition of the need for expansive data exchange and sharing capabilities:

> Rely on a national data sharing exchange platform, open data platform and other public infrastructure. Construct governance, public services, industrial development, technology research and development, and other fields of big data information databases. Support the implementation of national governance data applications. Integrate various types of social data platforms and data center resources. Create nationwide integrated service capabilities with reasonable layout and linkages. (State Council, 2017)

Despite the importance of data aggregation capabilities, it is unclear how much progress has been made with respect to the development of such capabilities. According to Liang et al. (2018), Chinese media reports often refer to major disagreements about data formatting, standardization, and interoperability between systems. In this respect, China is likely to confront many of the same technical and political issues confronting large-scale data aggregation efforts in the West, such as those associated with Open Government Data initiatives in the UK (https://data.gov.uk/) and US (https://www.data.gov/).

4.3.4 Data Processing

Data processing refers to the attempt to analyse social credit data with a view to calculating various metrics, most notably a social credit score. It is unclear how such scores will be calculated in China's SCS,[7] and existing SCSs (e.g., the local pilot projects discussed in Section 4.2.2) provide little in the way of insight into this issue. One possibility is that AI systems will be used to support the derivation of social credit scores. Support for this idea comes from the *New Generation Artificial Intelligence Development Plan*, which mentions the use of AI in the context of systems to "Promote social interaction and mutual trust" (State Council, 2017).

One of the issues surrounding the calculation of social credit scores is whether the relevant processing will centre on data that is specific to a given social actor (e.g., an individual citizen) or whether it will incorporate data pertaining to other actors. The relevant distinction here is between the processing of personal data on the one hand and interpersonal data on the other. Personal data, in this case, focuses on data about a specific individual—e.g., the things that a particular individual has done. Interpersonal data, by contrast, includes information about an individual's relationship to other individuals. Consider, for example, a state-of-affairs in which the social credit score of one individual (*A*) was affected (positively or negatively) by the social credit scores of other individuals within *A's* social network. The score assigned to *A* may, for example, be downgraded if it turns out that they are acquainted with individuals of questionable integrity. This is likely to act as a powerful *social network incentive* (see Pentland, 2014), in the sense that it acts to incentivize 'good' behaviour while simultaneously transfiguring the structure of existing social networks. Individuals with low scores may not only find themselves denied access to basic social services they may also find it harder to establish social connections. Given that a "need to belong" or a need for social connection is deemed to be a fundamental human motivation (see Baumeister & Leary, 1995), we would expect interpersonally-inflected forms of data processing to yield substantive effects on human behaviour, even in the absence of explicit reward or punishment schemes. Social connectivity (or isolation) is, in effect, its own reward (or

[7] Indeed, it is not even clear if such scores will be calculated at all. The SCP, for instance, makes no mention of social credit scores.

punishment).

4.4 (Dis)Honest Data

One of the issues confronting the development of the SCS is the reliability of credit evaluation processes. While the SCS strives to promote social trust and honesty, it will only work if the data upon which the evaluation process is based accurately reflects an individual's behaviour. If there is too much noise in the system, then credit scores are unlikely to tell us much about an individual's past behaviour or future conduct. Note that the goal of the SCS is to promote trustworthiness, honesty, sincerity, and so on in wider society. But the overall policy assumes that the data underpinning the SCS is itself honest or credible (i.e., reliable). This promotes a shift in thinking: from notions of social credit to issues of data credibility. From an epistemological perspective, a social credit score is intended to serve as a source of knowledge about an individual. In particular, it is intended to tell us something about how an individual will behave and thus the extent to which an individual can be trusted. If, however, the process by which the score is derived is an unreliable indicator of actual behaviour, then it is hard to see how the score can be expected to fulfil its intended purpose.

4.5 Incentivizing Vice

The SCS aims to promote honesty, sincerity, and integrity by establishing a system of rewards and punishments based around the notion of social credit scores. The system relies on the idea that 'good' behaviour is incentivized courtesy of the value attached to high social credit scores. Inasmuch as high social credit scores can be translated into rewards, then the assumption is that individuals will 'earn' such rewards by engaging in virtuous behaviours, i.e., behaviours that are deemed sincere, honest, and so on.

The problem with this logic is that valued commodities are just as likely to incentivize bad behaviours as they are good behaviours. In the effort to acquire monetary rewards, for example, some individuals will assume the role of an honest, hard-working citizen; others, however, will resort to more nefarious means of securing financial treasures. The point here is that there is no guarantee that individuals will always play by the rules when it comes to issues of social credit: a system designed to inculcate honesty may just as easily have the reverse effect, especially if the rules of the 'game' are not adequately enforced. Consider, for example, the Alipay system. According to Ohlberg et al. (2017), Alipay has given rise to black markets in which hackers are employed for the purpose of artificially boosting a user's Sesame Credit Score (e.g., fabricating records of educational attainment). Similar examples have arisen in respect of transport-related blacklists. Chorzempa et al. (2018) thus note that a robust system for circumventing China's travel blacklist has sprung up, with brokers offering to obtain train and air tickets for blacklisted individuals using alternative credentials.

The roll-out of the SCS could be accompanied by similar attempts to subvert the system as a means of protecting (or elevating) one's reputational status and thereby securing access to social resources. In this sense, we can see the 'duplicitous' nature of the incentives associated with the SCS: On the one hand, the SCS could promote virtuous behaviour via the promise of reward and the threat of punishment. On the other hand, however, the *same* set of rewards and punishments could incentivize attempts to sabotage, subvert, and manipulate the system, thereby leading to the forms of (dishonest) conduct that the system is intended to eliminate.

4.6 The Western View

Most Western commentators resolutely reject China's SCS. The SCS is commonly depicted as an instrument of government oppression, a tool for totalitarian surveillance, and the hallmark of an Orwellian society. While the fears associated with such dystopian rhetoric may be genuine, they are hardly ever based on an accurate understanding of what the SCS actually is. Given that there is no single SCS in effect at the time of writing (see Section 4.2), it is probably too early to make sweeping statements about what the SCS may become (and claims about what it actually is are clearly false).

A common mistake is to equate the SCS with a local pilot project, such as the Rongcheng SCS, or a commercial system, such as Sesame Credit. Given that pre-existing systems are a pretty mixed bunch as regards their aims, scope, reward schemes, and so on, such comparisons should be treated with some caution. While China's SCS may possess some of the features of existing SCSs, there is no hard evidence that the Chinese leadership is committed to any of these existing systems. Neither is there any evidence to suggest that the SCS will coalesce around an existing system, such as Sesame Credit. As noted in Section 4.2.3, Sesame Credit representatives have explicitly rejected claims they are secretly sharing credit-related data with government authorities. In any case, it is doubtful whether any of the existing SCSs deserve their dystopic reputation. Many of the systems are optional and seemingly innocuous. In fact, one of the problems confronting social scientific analyses of SCSs is that people are seldom aware of the existence of such systems. In a recent survey of Chinese public opinion, for example, Kostka (2019) notes that even in areas where a local pilot project was being run, only 11% of the residents reported being aware that they were a part of the project.

Another common mistake is to overplay the scope of the surveillance effort associated with the SCS. To be sure, there are plenty of surveillance-related efforts afoot in China,[8] and no doubt surveillance is a common feature of social life for Chinese citizens, at least in urban centres. Whether such systems can be tied to a government-run SCS is, however, another matter.

Facial recognition systems that identify (and sometimes fine) pedestrians who violate traffic violations are a neat example of a system that combines multiple technologies (e.g., AI, Internet connectivity, and personal data) for the purpose of shaping social behaviour. In the absence of any wider integration into an enduring nexus of personal data, however, the system is hardly more obnoxious than a speed camera on a UK High Street. The speed camera, of course, does not rely on facial recognition technologies to identify a culprit (as far as we aware), but it is not clear why that (by itself) should make a difference: setting aside the technical details, both cases involve a specific individual being identified courtesy of the use of image processing technologies and (network-mediated) access to bodies of personal data.

The status of the Internet as a surveillance technology for the SCS is also unclear. In particular, it is unclear to what extent China plans to monitor online behaviour (e.g., social media posts), evaluate such behaviour, and factor it into credit scores. While such forms of surveillance are a familiar feature of Western media narratives about the SCS, the 2014 SCP has very little to say about the Internet. In fact, the Internet is mentioned in only one paragraph of the SCP. In addition, what the SCP does say is focused on the use of the Internet as a tool for data sharing and the publication of blacklists. Here is the relevant paragraph in full:

> Forcefully move forward the construction of online sincerity, foster ideas of running the Internet according to the law and using the Internet in a sincere manner, progressively implement the online real-name system, perfect legal guarantees for the construction of online credit, forcefully move forward the construction of online

[8] Consider, for example, China's efforts to install a high-tech surveillance system in Xinjiang, an autonomous region in Northwest China (e.g., Buckley & Mozur, 2019).

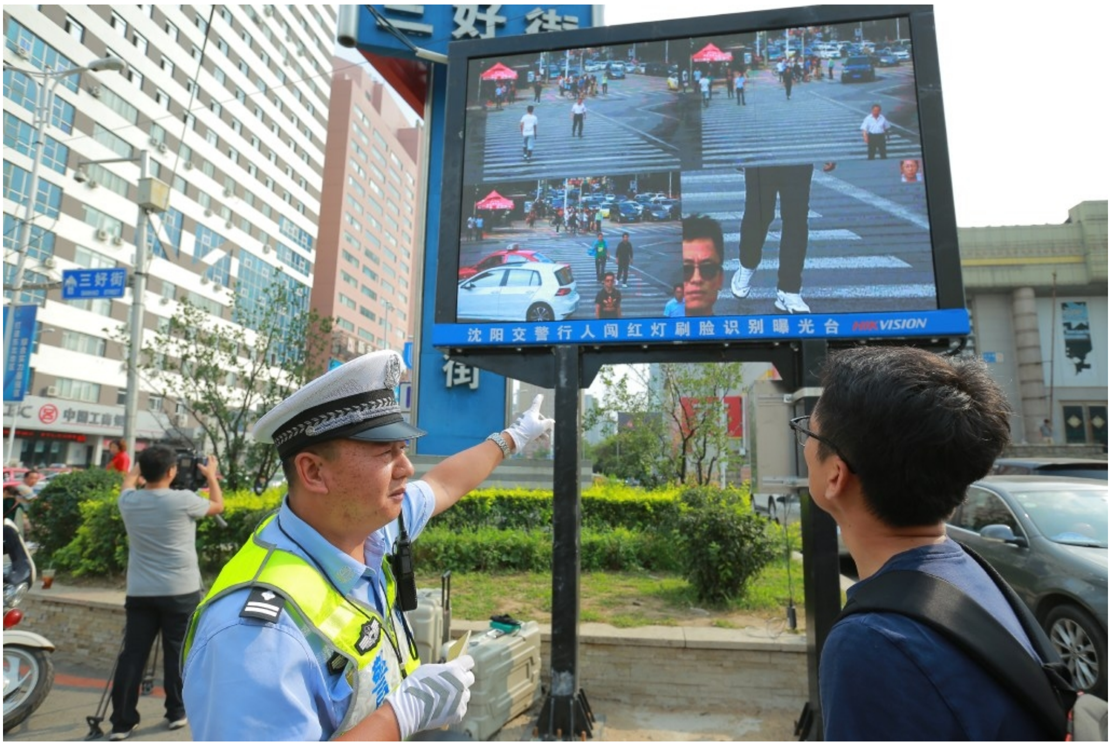


Figure 4.6: Use of facial recognition technology in Shenyang to prevent pedestrians from running a red light (source: Hiranand, 2018).

> credit supervision and management mechanisms. Establish online credit evaluation systems, evaluate the credit of the operational behaviour of Internet enterprises and the online behaviour of netizens, and record their credit rank. Establish network credit files covering Internet enterprises and individual netizens, vigorously move forward with the establishment of exchange and sharing mechanisms for online credit information and corresponding credit information in other areas, forcefully promote the broad application of online credit information in various areas of society. Establish online credit black list systems, list enterprises and individuals engaging in online swindles, rumour-mongering, infringement of other persons' lawful rights and interests and other grave acts of breaking trust online onto black lists, adopt measures against subjects listed on black lists including limitation of online conduct and barring sectoral access, and report them to corresponding departments for publication and exposure. (State Council, 2014)

In practice, of course, the Internet may play a greater role in the SCS than the length of this paragraph (relative to the larger SCP document) would suggest. Towards the end of the paragraph, we see "limitation of online conduct" mentioned as a potential punishment mechanism. Given what we already know about the scale of Internet penetration in China (see Section 3.4.2) and its incorporation into many aspects of Chinese social life (see Section 3.4.3), constraints on Internet use may be more consequential than we might otherwise expect.

Although the extent to which online behaviour will be incorporated in China's SCS remains a moot point, it is clear that the Internet provides ample opportunities for social surveillance. This stems from the fact that many forms of social activity now involve the Internet. Earlier, we referred to this "social entrenchment" (see Section 3.4.3). One of the implications of social entrenchment is that it invites us to reconsider the traditional dichotomy between online and offline behaviour. Inasmuch as we accept the idea that the Internet is an increasingly important part of social life—part of the physical, mechanistic fabric that makes all manner of social activities materially possible—so the erstwhile crisp distinctions between the "online" and "offline" begin to blur. This point is made by Floridi (2014) who argues that the online/offline dichotomy ought to be replaced by a unitary concept, which he refers to as "onlife." The point here is that as the Internet becomes an increasingly important part of social life, so it becomes ever-harder to distinguish between its social and non-social roles. If, for example, the Internet is a component of everyday social activities, then it is nicely poised to play a role in the surveillance of those activities. The Internet may, in short, function as a 'window' that sheds light on many forms of social activity, including those that private citizens, commercial organizations, and government officials would prefer to keep hidden from prying eyes. The upshot is that we should not assume that the surveillance potential of the Internet is limited to the detection of anti-government sentiments on popular social media platforms.[9] In all likelihood, the Internet's status as a surveillance technology will expand as it becomes an ever-more important part of the material mix that shapes the nature of social reality.

4.7 Western Parallels

While the SCS has been discussed as a Chinese phenomenon—something that is specific to China—it should be clear that many of the features of the SCS are not entirely alien to a Western audience. At a general level, the SCS seeks to adapt the credit scores that are commonly encountered in the financial domain and apply them to a broader array of social activities. In this sense, it is worth asking ourselves what it is that motivates the rejection of the SCS by Western commentators. This question is important, not just because it directs attention to Western counterparts of the SCS, but also because it improves our understanding of the parameters for SCS design. Such an understanding may be useful when it comes to evaluating the extent to which China's SCS could be adapted for use in other countries.

As noted, the SCS bears similarities with financial credit scoring systems, which are a common enough feature of Western societies. In both cases, we see how access to certain kinds of resources (financial or otherwise) are constrained courtesy of a computed metric—a credit score—that is calculated with respect to bodies of personal data. Relative to this abstract characterization, we can discern some obvious points of divergence between financial and social credit systems. These include:

- The kinds of (personal) data that are used to compute a credit score.
- The kinds of algorithms that are used to compute a credit score.
- The kinds of resources whose accessibility is affected by a credit score.

In respect of the first point—the kinds of data used to compute a credit score—it is natural to assume that the scope of financial credit systems is limited as compared to their social credit counterparts. In evaluating someone's suitability for a loan, for example, many financial credit systems limit their attention to data of a financial nature, for example, an individual's credit history, income, patterns of consumption, and so on.[10] As some have noted, however, the credit scores formulated by some financial companies have shown signs of going beyond the bounds of pecuniary variables (Arya, Eckel, & Wichman, 2013). Wong and Dobson (2019) document a number of cases where the scope of financial credit scoring appears to extend beyond the

[9] Although the detection of anti-government sentiments, as well as other challenges to government authority, no doubt remains a prominent area of concern for the Chinese leadership (see King et al., 2013).

[10] In the US, the FICO system uses information about types of credit used, payment history, length of credit history and the amounts owed in order to calculate one's total FICO score (Hurley & Adebayo, 2016).

realms of purely financial data. In the US, for example, they note that a financial services company, Affirm, has moved away from traditional credit reporting to the scanning of one's profile on social media platforms like Facebook. The goal, they suggest, is to evaluate whether someone is deemed worthy of a loan. Wong and Dobson (2019) also cite the case of Lodex, an Australian financial services company, that purportedly predicts the likelihood of loan repayment via an analysis of an individual's smartphone usage and emails. Other evidence in support of a broadening of scope for financial credit evaluations is readily available. Síthigh and Siems (2019) cite the example of Lenddo, a Singapore-based lender, which, purports to use "nontraditional data…to economically empower the emerging middle class" (p. 16) Such data is deemed to include information from social media systems. Another company, Tala, which operates in Africa and Asia, is said to utilize "up to 10,000 'data points' such as social media and smartphone use, in order to create a new type of credit score to the advantage of lower-income customers" (Síthigh & Siems, 2019, p. 16).

As evidenced by this quote about Tala's commercial objectives, it would be a mistake to assume that by broadening the scope of credit scoring systems, financial companies are in any way jeopardizing the financial interests of the individuals they serve. While access to detailed credit histories may be a familiar feature of Western, developed nations, such data may not be readily available in developing nations. In this case, the widespread availability of social media data may serve as a proxy for financial data, helping to reduce a lender's uncertainty about individuals with limited financial histories, and thus improving the access that economically disadvantaged individuals have to finance (see Wei, Yildirim, Van den Bulte, & Dellarocas, 2016, for a useful overview).

There is, of course, a subtle shift in emphasis here. In the case of conventional financial credit scoring, a credit score is intended to serve as an estimate of an individual's capacity (and propensity) to repay a sum of money. In the absence of personal financial records, however, the credit score is apt to tap into a broader array of psychosocial characteristics, including aspects of an individual's character, conduct, and their position in social networks. This is not to say that the goal of credit scoring differs across the two cases. In general, it seems sensible to assume that what lenders are interested in is a measure of (financial) trustworthiness. That is to say, each lender is interested in the extent to which an individual can be trusted to fulfil their repayment obligations. Issues of financial standing are clearly not irrelevant here: in order to repay a loan, one must have the financial wherewithal to repay it. Nevertheless, a mere capacity to repay a loan is not sufficient; one also needs to be able to trust an individual to repay a loan if they have the means to do so. Issues of trust are thus a common feature of financial credit scoring systems, and, inasmuch as non-financial data can be used to more accurately assess an individual's trustworthiness, there is ample incentive for financial companies to incorporate such data into their credit score calculations.

It should not be assumed that this broadening of scope in financial credit scores is a uniquely modern-day phenomenon, brought on by the advent of the Internet and the burgeoning popularity of social media systems. As noted by Síthigh and Siems (2019):

> The history of credit registries and scores identifies a long-running 'character' dimension, including the desired impact upon consumer behaviour, the framing of a good credit score as a moral virtue, the use of data (via informants or otherwise) on personal character, and attempts to incorporate factors such as 'honesty' and 'clean living' into scoring. (Síthigh & Siems, 2019, p. 33)

Síthigh and Siems (2019) base their analysis on work by Lauer (2017), which charts the history of US credit reporting bureaus. As part of this analysis, Lauer highlights the value-laden nature of early credit reporting systems, which emphasized the importance of conduct, character, and moral virtue as the cornerstone of (financial) credit evaluations. Relative to this historical backdrop, many of the features of China's SCS (e.g., the scope of the credit evaluation process, the value-laden nature of social credit scores, and the more general valorization of culturally-specific forms of virtue) begin to look a lot like the features of systems that were prevalent in the early days of consumer capitalism.

A second distinction between financial and social credit systems relates to the nature of the algorithm that is used to compute credit scores (see above). Comparative analyses in this area are complicated by the fact that such algorithms are seldom available for public scrutiny. As noted above, it is unclear what algorithms will be used to compute credit scores in the context of China's SCS (see Section 4.3.4). In addition, most of the local pilot projects currently operating in China do not provide details of their credit scoring algorithms. For the time being, then, the differences between credit scoring algorithms (in respect of financial and social credit systems) remains obscure.

There are likely to be many features of credit scoring algorithms that influence the overall effectiveness and cultural acceptability of SCSs. A particularly important source of inspiration, in this respect, stems from the various forms of reputation rating system that have emerged on the Internet. Such systems are unlikely to require much in the way of an introduction. They are a common (almost ubiquitous) feature of systems that are intended to support some form of social or economic exchange relation between unknown (and sometimes anonymous) social actors. Prominent examples of such systems include the likes of eBay, Uber, and Airbnb. The aim, in each of these cases, is to reduce the uncertainty that the recipients of some service have with respect to the vendor or provider of a service. A history of prior endorsements by the purchasers of products on eBay, for example, conveys information to future purchasers about the extent to which a vendor can be trusted to fulfil purchase orders.

An interesting feature of many of these systems is the presence of bidirectional rating schemes. Thus, it is not just the reputation (or trustworthiness) of the vendor that is being evaluated in exchange scenarios, the customer is also subject to evaluation. Often this customer-oriented evaluation relates to issues of payment, but it may, at times, appeal to other properties of the individual or the exchange relation. One example of a bidirectional rating scheme comes in the form of ridesharing services such as Uber, where both drivers and passengers have the opportunity to rate each other. A driver with a comparatively low score may be removed from the platform or restricted to providing a limited range of services. By contrast, a passenger is more likely to be picked up if they have a high rating from past drivers (see Síthigh & Siems, 2019).

From a social credit perspective, there is an important difference between online reputation systems and China's SCS. China's proposal centres on a credit evaluation system that is managed by a central government authority and the resulting ratings thus reflect the values, beliefs, and precepts of the CCP. This contrasts with the much more decentralized approach to scoring that is adopted by online reputation systems. In the case of systems like Uber, eBay, and Airbnb, individual scores are ultimately determined by the information provided by other users, based on their experience of particular social exchanges. There is no reason to assume that such ratings do not reflect value-laden judgements; neither is there any reason to assume that every rating targets exactly the same features of the transactional context: some ratings, for example, may reflect performance, while others may target factors such as courtesy, friendliness, or effort. What is crucial here is the way in which some score (a social credit score or a reputation ranking) is derived as a result of the individual decisions made by individual actors. The scoring process is, if you like, delegated to the community of actors that must ultimately live with the consequences of their decisions. It is then up to the community (or at larger scales, entire societies) to determine where to locate the boundary between vice and virtue—between the sort of things that are worth promoting and the sorts of things that are worth punishing. In all likelihood, it is this *locus of* *control issue* that lies at the heart of much of the Western disquiet about the SCS. In contrast to contemporary online reputation rating systems, China's vision for social credit is one in which decisions about what constitutes 'good' and 'bad' are made by a powerful elite that has a vested interest in safeguarding its political authority. The risk, of course, is that in such situations the notions of vice and virtue will be ones that are defined relative to whatever it is that enables such an elite to maintain its grip on power.

What, finally, of the third differentiator between financial and social credit scoring systems—the resources whose accessibility is affected by a credit score? Given the objectives of China's SCS, it is likely that its effects will be felt across many sectors of society. In this sense, the consequences of a low social credit score, in the context of the SCS, will have implications that go beyond those associated with a low financial credit score. At the same time, however, it should be remembered that the effects of financial credit scores are not constrained to some purely economic realm that is utterly divorced from other aspects of social reality. Inasmuch as access to financial credit affects one's overall life prospects, such as determining where one can live and what education one can receive, it is far from clear that the consequences of a low social credit score in China are any more detrimental to one's socio-economic well-being than is a low financial credit score in the US.

Having said that, there is no doubt an important difference between financial and social credit scores as regards the way in which they are used to restrict or enable access to resources. This difference stems from China's vision for a so-called Joint Punishment System (see State Council, 2016b). A key feature of this system is that transgressions in one area of social activity are punished via the imposition of constraints in other areas. (Imagine, for the sake of example, a state-of-affairs in which a conviction for a road traffic violation—e.g., a speeding offence—affects one's ability to buy a train ticket.) China's plans for the Joint Punishment System centre on the use of blacklists, where information about inappropriate conduct is made available to a variety of social actors. At present, it is unclear who will have access to these blacklists. Evidently, the government intends for the blacklists to be used across different government departments, but it remains unclear whether the blacklists will be used to guide decision-making in non-governmental contexts. Another unknown relates to the scope of the blacklists themselves. One possibility is that the government will seek to maintain multiple blacklists, each of which targets a specific sphere of social activity. Such blacklists would resemble those used in Western democracies, such as "no-fly lists" in the US and "football banning orders" in the UK (see Chorzempa et al., 2018). On the other hand, the Chinese government may seek to maintain a limited number of blacklists that pool data from different social domains (or spheres of social activity).

In Western societies, blacklists (financial or otherwise) typically have both limited scope and limited distribution. For the most part, one's financial credit rating does not affect one's access to government services or one's ability to book a flight. Similarly, financial credit scores are not available to everyone, although the list of organizations that can access such information is sometimes extensive. According to the credit reference agency Experian, the companies that can access credit report data include banks, mortgage providers, creditors and lenders, potential employers, utility and service companies, mobile phone companies, letting agents and landlords, debt collection agencies, government agencies, and insurance companies.[11]

In contrast to financial credit scores, the ratings generated by online reputation systems are widely available. Even if they are not 'publicly available', in the sense of being available to everyone on the Internet, they are still visible to registered members of a particular platform. For the time being, the scope of such ratings or 'scores' tends to be somewhat limited. One's rating on Uber, for instance, has very little impact on one's rating in other systems. The extent

[11] see https://www.experian.co.uk/consumer/guides/who-has-access.html.

to which this will always be the case is unclear. As noted by Síthigh and Siems (2019), there are some obvious advantages associated with the capacity to merge reputational information across different platforms. This includes an opportunity to alleviate the ‘barriers to entry’ (see Section 2.6) faced by newcomers who lack a reputational history on a specific platform.

The capacity to integrate data across disparate social media systems has been a recent focus of attention within the academic community. According to Hendler and Berners-Lee (2010), one of the limitations of contemporary social media systems is that they operate in isolation from other systems. Hendler and Berners-Lee thus call for greater integration and cross-communication between these systems, which, they suggest, will provide new opportunities to resolve pressing social problems. Hendler and Berners-Lee are undoubtedly right to note the advantages associated with large-scale forms of data integration. But note that such forms of integration also yield a capacity to link reputational information (as well as other kinds of personal data). This comes with an attendant risk. The danger is that one’s reputational shortcomings in the context of a single system are apt to influence one’s reputational standing across a multiplicity of different systems. At this point, we seem to have reached a state-of-affairs that is not all that different from the vision portrayed in the Joint Punishment System—a system in which transgressions in one sphere of social activity has implications for one’s ability to access resources in many other areas.

5 — The Value of Data

> *Data-driven innovation holds the keys to addressing some of the most significant challenges confronting modern Britain, whether that is tackling congestion and improving air quality in our cities, developing ground-breaking diagnosis systems to support our NHS, or making our businesses more productive. The UK's strengths in cutting-edge research and the intangible economy make it well-placed to be a world leader, and estimates suggest that data-driven technologies will contribute over £60 billion per year to the UK economy by 2020.*
>
> —HM Treasury (2018)

What are the implications of the foregoing analysis for our understanding of issues relating to data value and the strategic importance of data in a geopolitical context? Many nations now regard data as a valuable resource. This is reflected in the above quotation, which represents the opening statement of a recent UK Treasury report on the economic value of data. Here, the value of data is judged relative to a set of economic and social benefits, many of which stem from data's capacity to enhance productivity and improve the quality of decisions. This applies just as much to the realms of government as it does the world of business. Courtesy of their access to large-scale bodies of social data, national and regional governments hope to improve their understanding and awareness of society and thereby improve the effectiveness of governmental processes.

Data also provides opportunities for insight and innovation. In contrast to the notions of social and economic value, this is what might be called the epistemic value of data. The capacity of data to yield advances in our scientific understanding of the world is, of course, sufficiently obvious as to warrant little in the way of further discussion. Recently, however, data has assumed a new significance with an explosion of interest in machine learning and a growing appreciation of the economic, social, and military potential of AI capabilities. In this respect, the value of data inheres in its ability to support the development (and sometimes drive the operation) of new forms of AI system. Such value is widely appreciated by AI practitioners and theorists. Hall and Pesenti (2017), for example, note the importance of data as part of their review into the UK AI industry. They provide a number of recommendations that are specifically intended to bolster the UK's access to data for the purpose of developing innovative AI solutions. These include the provision of frameworks to support data sharing, the use of representational formalisms to facilitate data exploitation, and the development of data publication and data description methods that provide information about the potential use of data for AI purposes.

The present section discusses a number of issues that relate to the value of data (specifically, personal data). In addition to the links between data and the development of AI capabilities, we also explore some of the links between value, trust, and social governance.

5.1 Artificial Intelligence and Data Value

Given the recent burgeoning of interest in machine learning, it is perhaps not surprising that our intuitions about the value of data should have shifted in recent years. It is typically assumed that data is valuable when it comes to the development of AI capabilities. As noted in Section 3.8, however, assessing the value of data in an AI context is not a straightforward task. We have seen, for example, that not all forms of AI innovation require access to human-generated data (see Section 3.8.3). This directs attention to the different kinds of data that might be used to support the acquisition of intelligent capabilities. Given that not all capabilities require access to bodies of personal data, it is perhaps a mistake to regard a nation's AI potential as directly related to the amount of personal data it has at its disposal. No doubt personal data is invaluable for certain purposes. One cannot, for example, rely on synthetic data to train a machine learning system if one has no understanding of the various causal forces and factors that conspire to generate the data in the first place. Nevertheless, it is not clear that *all* forms of AI innovation—and thus a nation's standing on the international stage—require access to personal data.

What is it, then, that dictates the value of data from an AI perspective? Undoubtedly, part of the answer to that question relates to the capacity of particular bodies of data to support the acquisition of certain kinds of capability. If, for example, the aim is to detect the presence of objects of a given type (e.g., firearms) in 2D images, then the training data will need to include objects of the relevant type. This, however, is not the end of the story, since issues of scope and volume are also relevant to data valorization. In respect of scope, for example, it is important to know how the training data (the sample) was obtained from the target population. Given that any form of sampling bias is apt to affect the performance of a system, it is important to understand how the statistical properties of a dataset relate to the features of the domain in which an AI system is to be deployed. As a simplistic example: If the goal is to estimate the mean height of adult individuals in a population, it will not help us if the sample includes a preponderance of data from men, and men are, on average, taller than women.

Issues of data volume (or data quantity) are also apt to be important. In general, the larger a sample, the better the estimate of population parameters. But the relationship between volume and value is not necessarily linear. A small dataset may be of no value whatsoever for the purpose of training a machine learning system. Similarly, beyond a certain point, there may be little value in securing access to additional data. Much depends on how accurately we want to model the domain in which an AI system is expected to operate. Returning to our simple case of estimating adult heights, a (random) sample of 1000 individuals will provide a better estimate of the population mean than will a (random) sample of 100 individuals. But if there is a cost associated with the acquisition and processing of sample data—where such costs involve issues of privacy violation and computational efficiency, in addition to the more traditional financial costs—then a question arises as to when our sample size is 'good enough'. That is to say, at what point do we have enough data to deliver an acceptable level of performance for a given application?

In answering this question, it will be important to consider the way in which data can be valorized relative to specific AI objectives. This is likely to require a concerted effort to understand how the features of a dataset are apt to support the acquisition of specific capabilities. Arguably, such assessments will require access to information about how specific bodies of data were collected, what such data represents, and (perhaps) what AI capabilities might best be supported by the data. Such metadata could be captured in the form of *data manifests* that exploit emerging standards for the representation of provenance-related information (see Moreau, 2010). There is also ample opportunity here for the use of AI techniques as part of the data valorization process. The upshot is a vision of what might be called *intelligent data valorization*, namely, the use of AI systems to evaluate the use of data for specific training purposes. Such systems could perhaps play an important role in future data trading activities.

5.2 The Importance of Trust

It is widely assumed that China is 'ahead of the game' when it comes to data access (see Section 3.4.1). This has obvious implications for China's standing on the geopolitical stage. Inasmuch as we accept the idea that AI capabilities are driven by data access, then it seems that China's position as a world-leader in AI innovation is assured. At the same time, however, there are some obvious constraints on a nation's ability to leverage the value of its domestic data for AI purposes. One such constraint relates to issues of computational power. Machine learning is a resource-intensive process that comes with substantive computational overheads. In addition, AI systems are often limited by the nature of the semiconductor technology that is used to realize their functionality. In this sense, access to semiconductor technology is a key feature of a nation's AI potential (see Section 3.6).

China's government has clearly worked to resolve issues that threaten to stymie its AI ambitions (regardless of whether or not it has been successful). But it is not clear that all its actions are consistent with its ambitions to become a global AI superpower. When it comes to the collection and use of personal data, for example, China has shown signs of moving towards a regulatory framework that resembles that adopted by the EU (see Section 5.6). As was mentioned in Section 3.4.4, China's actions in this space are probably best interpreted as a trust promotion exercise, i.e., as a means of promoting public trust in companies that collect and curate personal data.

Issues of trust also lie at the heart of China's approach to AI governance. In 2018, China produced the *White Paper on Artificial Intelligence Standardization* (China Electronics Standardization Institute, 2018), which highlights the importance of public trust in AI technologies: people must trust AI technologies, it suggests, otherwise such technologies will not be accepted and their socio-economic benefits will be out of reach. The white paper goes on to note some of the risks posed by the development of AI systems, including the capacity of AI systems to infer information about an individual and thereby violate expectations about the sorts of (personal) data that are available to such systems. The danger here is that AI systems may evolve in such a way as to undermine the foundation of public trust that contributed to their original development. That is to say, once AI systems are able to make inferences about information that goes beyond what individuals have agreed to disclose, then it is unclear whether individuals will agree to provide the sort of data that makes these inferences possible.

This highlights the importance of trust to AI development. Inasmuch as AI systems benefit from access to personal data, then public trust is apt to serve as an important constraint on AI development. In the absence of such trust, the public may resist efforts to collect the bodies of data that are necessary for at least some forms of AI research. Such a vision portrays progress in AI as something of a large-scale (i.e., society-wide) cooperative exercise—one that benefits from the public's support for such efforts. The absence of such support is likely to serve as a significant constraint on AI innovation, and this is so regardless of whatever technological and intellectual resources a country has at its disposal.

5.3 Social Governance and Personal Data

Issues of trust also lie at the heart of China's approach to social governance. In particular, China's SCS is intended to document trust-breaking behaviours and thereby provide incentives for honest and sincere forms of social conduct. Given that an effective SCS is likely to require access to bodies of personal data, we can see how China's proposals in this area provide an additional incentive for the acquisition of personal data, one that is, in this case, justified via an appeal to issues of trust.

China's proposals in respect of the SCS are, of course, predicated on an ability to collect and process personal data, especially at large scale. But, in addition to being enabled by a vibrant online data ecology, the SCS is also likely to contribute to an *expansion* of that ecology. By using personal data as a tool for social governance, China's leadership provides incentives for data collection and improvements in its data infrastructure (e.g., improvements in data integration and sharing mechanisms). All of this is done in the name of trust. That is to say, the purpose of the SCS is to support trust-related evaluations and promote the adoption of trustworthy behaviours. No doubt the SCS plays an important political role, in the sense that it reflects the ideological idiosyncrasies of the Chinese leadership, but this should not detract from the trust-related role of the SCS. Neither should we overlook the status of the SCS as a complex sociotechnical system that is enabled by a technological/data ecosystem on the one hand, and which also acts as an enabler for changes to this ecosystem on the other.

5.4 Data Credit Systems

As noted above, issues of trust are an important element of China's SCS. By monitoring the behaviour of social actors, China hopes to provide information about trust-breaking behaviours and thereby provide incentives for certain forms of social conduct. The use of the term "social actors" (as opposed to individual citizens) is important here, for the SCS applies just as much to commercial organizations as it does to individuals. In addition, the reach of the SCS is not simply confined to Chinese companies, it also applies to foreign companies that operate within China. According to a recent report titled *The Digital Hand: How China's Corporate Social Credit System Conditions Market Actors* (European Union Chamber of Commerce in China & Sinolytics, 2019), foreign companies will be subject to continual monitoring, with inappropriate behaviour affecting a company's credit score and thus their ability to operate in the Chinese marketplace:

> The Corporate SCS is part of a fundamental shift in China's market access regime. It is the government's tool to ensure that only companies that prove to be trustworthy can access and participate in the Chinese market. The Chinese Government aims to implement a 'survival of the fittest' mechanism, but with a new definition of 'the fittest'. Under China's new market access regime, the fittest will be those deemed 'trustworthy', with 'trustworthiness' also being defined by the Chinese state, and monitored and enforced through the Corporate SCS. (European Union Chamber of Commerce in China & Sinolytics, 2019, p. 7)

For a variety of reasons, China's SCS proposals are unlikely to receive a sympathetic hearing among a Western audience. This does not mean, however, that we should abandon the whole idea of credit scores as a means of shaping social behaviour. China's SCS provides us with a single instance of what is likely to be a much larger space of design possibilities. Much opposition to China's SCS, and the quite palpable unease it causes in the Western media, is no doubt rooted in the way social credit scores embody the value-laden judgements of a political elite rather than an outright rejection of the bedrock idea that social actors ought to be rewarded (or punished) for compliance (or non-compliance) with normative standards of social conduct. China's SCS provides us with a vision in which decisions about what constitutes 'good' behaviour are made by a centralized authority. It is also a system in which credit evaluations are likely to be made by systems and procedures that are subject to government approval. This, however, is not the only way in which SCSs could be implemented. In contrast to China's more centralized vision, for example, we can imagine systems in which decisions about what constitutes good behaviour are delegated to society-at-large.

Similarly, there is no reason why the computation of social credit scores needs to be performed by a centralized, government-controlled computational system that excludes the possibility of social participation. Instead, we can imagine situations where the social actors themselves participate in the computational processes by which social credit scores are derived. Such forms of participatory involvement (or social computation) are exemplified by any number of online reputation systems, including those implemented by the likes of eBay, Uber, and Airbnb.

One area for future research is thus to explore the design space for SCSs and assess the acceptability of particular systems for different geopolitical constituencies. As part of this work, it will be useful to examine the extent to which the principles of social credit can be applied to the trust relationships that emerge in respect of personal data. Perhaps, for example, we can imagine a *data credit system*, in which social actors (most notably, companies) are evaluated with respect to their data collection and data handling conduct. The resulting data credit scores might then work to ensure that companies operate in accordance with the values that are relevant to the communities that are the focus of their data collection efforts. Note that such a vision need not presuppose any understanding of what constitutes 'good' behaviour' or how the notion of trustworthiness ought to be defined. Rather than resort to regulatory fiat as a means of imposing such definitions on a community, it may be better to view the community as a source of solutions to normative problems, i.e., to view appropriate forms of conduct (and, indeed, what constitutes appropriate conduct) as an emergent feature of systems in which social transactions are coordinated with respect to socially-derived measures of reputational standing.

5.5 The Value of Data Flows

In recent years, China's approach to data governance has been characterized as a form of data protectionism or data localism. In particular, China's 2017 *Cybersecurity Law* imposes constraints on the storage and transfer of data that is produced within China. According to Article 37 of the *Cybersecurity Law*:

> Critical information infrastructure operators that gather or produce personal information [i.e., personal data] or important data during operations within the mainland territory of the People's Republic of China, shall store it within mainland China. Where due to business requirements it is truly necessary to provide it outside the mainland, they shall follow the measures jointly formulated by the State cybersecurity and informatization departments and the relevant departments of the State Council to conduct a security assessment; where laws and administrative regulations provide otherwise, follow those provisions. (Standing Committee of the National People's Congress, 2017)

In all likelihood, such a policy reflects China's growing appreciation of the strategic importance of data. According to Sacks et al. (2017), "the Chinese government believes that data is on par with other natural resources, such as oil and gas, and requires a protection system that provides the government with insight into what types of data are flowing across borders." The upshot, they suggest, is that “Beijing’s approach...is more expansive than that of the European Union, which views data protection primarily through the lens of user privacy.”

There is, however, evidence that China recognizes the importance of cross-border data flows when it comes to economic growth and development. Dr Yanging Hong, one of the leading figures associated with China’s data protection policy, notes that:

> A fundamental consensus has emerged today that data naturally flows across national borders, that data flows produce value, and that data flows can lead to flows of technology, capital, and talent. Therefore, data flows are the norm, and circumstances where flows are limited are the exception. (cited in Sacks et al., 2017)

Relative to the present analysis, Dr Hong’s remarks are important, for they highlight the way in which one kind of flow (i.e., the flow of data) can provide the basis for flows of other kinds (e.g., flows of technology, capital, and talent). This speaks to a number of issues that were discussed in Section 3. They include the importance of intellectual capital as an enabling factor for China’s AI ambitions (see Section 3.5.1) and the role of migratory flows (especially with Taiwan) in shaping China’s domestic semiconductor industry (see Section 3.6.8). Further research in this area should seek to understand the value of data flows, not just in terms of their capacity to drive AI innovation, but also in the sense intimated at by Dr Hong, namely, the capacity of data flows to facilitate the flow of other kinds of resources.

5.6 Data Governance and The Four Internets

Section 3.4.4 briefly discussed China’s emerging regulatory framework in respect of personal data. While China is sometimes seen to have a relatively lax regulatory regime that affords better commercial opportunities for the exploitation of data, China’s regulatory environment is in a state of flux, and the Chinese government’s actions in this area indicate that it is ready to legislate in favour of tighter controls over the acquisition, use, and sharing of personal data. The *Personal Information Security Specification*, which took effect in May 2018, outlines the Chinese government’s position with respect to personal data. It lays out granular guidelines for user consent and specifies how personal data (called “personal information” in the specification) should be collected, used, and shared (see Shi, Sacks, Chen, & Webster, 2019). A second publication, released in 2019, introduced a number of amendments to the original *Personal Information Security Specification* (see Luo, 2019). These amendments imposed constraints on how personal data could be collected (via the imposition of constraints on so-called “bundled consent”). They also introduced additional constraints pertaining to data sharing. In particular, when a personal data controller allows third parties to access personal data through their products and services, they are required to establish contracts with third parties, notify personal data subjects that certain products and services are provided by third parties, ensure that third parties obtain consent from personal data subjects, and so on.

Although it is probably a mistake to equate China’s *Personal Information Security Specification* with the EU GDPR,[12] there are some obvious similarities between the two policy frameworks, at least when it comes to the language used in the relevant policy documents.[13] This raises questions about the extent to which China’s approach to personal data regulation ought

[12] See Gentle (2018) and Sacks (2018a) for comparisons between the EU GDPR and China’s *Personal Information Security Specification*.

[13] This is not to say that China’s approach to personal data protection is the same as that enshrined in the GDPR. There may, for example, be significant differences in the way particular policies are implemented. This is what might be called the “implementation gap.” The implementation gap is important, for it threatens to undermine efforts to compare data governance frameworks (see Section A.2.2).

Table 5.1: Chinese counterparts of US Internet services. Note how many of the Chinese services are owned by one of three companies, namely, Baidu, Alibaba, and Tencent.

US	**China**
Google Search	Baidu Search (Baidu)
eBay	Taobao (Alibaba)
Amazon	Tmall (Alibaba)
Amazon	JD
Amazon	RenRen
Facebook, iMessage, Slack	WeChat (Tencent)
PayPal	Alipay (Alibaba)
Uber	DiDi
Twitter	Sina Weibo (Alibaba 30%)
Twitter	Fanfou
YouTube (Google)	Youku (Alibaba)
YouTube (Google)	Tencent Video (Tencent)
Netflix	iQiyi (Baidu)
iTunes	QQ Music (Tencent)
Tinder	Tantan
Twitch (Amazon)	DouYu (Tencent 37%)
LinkedIn (Microsoft)	MaiMai

to be regarded as markedly different from the EU's approach. It also raises questions about the extent to which China's capacity to leverage its personal data ecosystem (e.g., for commercial and/or technological purposes) is hampered by its regulatory framework and whether the EU is in any way disadvantaged (relative to China) by its adoption of data protection laws. In respect of this latter issue, it might be thought that China and the EU are on a par: If it is indeed the case that China and the EU are converging on a common approach to personal data regulation, then is there any reason to think that one is more or less hampered (or helped) relative to the other on the international stage?[14]

In answering this question, we suggest it is important to consider the differences in the commercial Internet landscape between the EU and China. (Some of these differences were discussed in Section 3.4.6.) In particular, we suggest that despite the ostensible similarity of the two data protection regimes, the impact of these regulatory frameworks are not the same in the two geopolitical contexts. In other words, similarity of regulatory framework does not imply similarity of impact: the same regulatory framework could have differential effects on a country's capacity to leverage the value of data assets according to the specific socio-political and socio-economic context in which the framework is deployed.

Given the differences in the commercial landscape between China and the EU, we suggest that China's regulatory framework may not have the same impact as the EU GDPR, regardless of whatever similarities exist between the two frameworks. This is because China's Internet is dominated by a limited number of large commercial companies (e.g., Baidu, Alibaba, and Tencent) (see Table 5.1). This means that each company has a greater slice of the domestic data pie and is therefore probably less likely to be affected by regulatory constraints that impede the

[14] Further information about the similarities and differences between China and the EU in respect of personal data regulation is likely to be revealed with the release of China's *Personal Information Protection Law* and the *Data Security Law*. Both of these laws are expected to come into effect in 2020.

flow of data across organizational boundaries. This *differential impact hypothesis* is reflected in comments by O'Hara and Hall (2018) in regard to the Four Internet model discussed in Section 2:

> Where China and the United States are each large centralized markets, enabling the gathering of giant quantities of data to fuel their algorithms, Europe is more fragmented, both in terms of markets and in terms of the dominant tech companies, and this decentralization is exacerbated by the GDPR's stern regulation of data sharing. (O'Hara & Hall, 2018, p. 7)

The main implication of the differential impact hypothesis is a call for greater awareness about the consequences of government regulation relative to the way in which the Internet has evolved in different geopolitical regions. When it comes to issues of value, we should not assume that a common policy framework will be suitable for all nation states. Instead, data governance policies will need to be adapted to the idiosyncratic features of the local landscape. Given that differences in this landscape are, at least in part, a response to the political and ideological characteristics of specific regional and national governments, we can begin to see the beginnings of a self-reinforcing cycle of regulatory intervention that threatens to accelerate the demise of a common global Internet model. According to our discussion of the four Internets, different kinds of Internet model arise in response to the disparate interests of different geopolitical players. According to the differential impact hypothesis, such differences merely accentuate the need for different policy frameworks when it comes to attempts to press maximal strategic benefit from specific Internet models. The upshot is a vision in which the differentiation (and, perhaps ultimately, the fragmentation) of the Internet is the result of geopolitical differences and also the means by which those geopolitical differences might be accentuated.

6 — Conclusion

The present review is not intended as an exhaustive analysis of China’s approach to data, social, cyber, and AI governance. Nevertheless, it does provide some insight into issues that are likely to inform the shape of future research efforts. Here we outline a number of areas for future work that draw on the material presented in previous sections. A complete listing of research areas is provided in Appendix A.

China’s plans for a SCS are arguably a reflection of its authoritarian approach to social governance. At the same time, however, the SCS stands to benefit China’s AI ambitions by providing incentives for the collection of personal data and creating opportunities for AI development. In some sense, then, the SCS is a system with a dual purpose: on the one hand, it serves a political purpose, helping to preserve the power and stability of the CCP; on the other hand, it serves a technological purpose, helping to provide opportunities for technological growth and innovation. Given the West’s rejection of the SCS, it is difficult to see how any of the technological advantages of the SCS could be exploited by Western countries. This raises a more general issue about the extent to which China’s AI ambitions are facilitated by the features of Chinese society. Does China enjoy a strategic advantage, relative to other countries, simply because of the nature of its culture, history, and political system? Does China’s place in the geopolitical landscape provide a means by which it is able to dominate that landscape? Such advantages may stem from more than the roll-out of the SCS. In Section 3.4.7, for example, we saw how the Chinese leadership is able to promote the adoption of new technologies (e.g., driverless cars) in ways that might be difficult or impossible in the West. All this directs attention to the way in which China enjoys something of a *socio-political advantage* when it comes to AI innovation. To what extent do the features of Chinese society (its history, culture, and political system) support China’s efforts to become a global AI superpower?

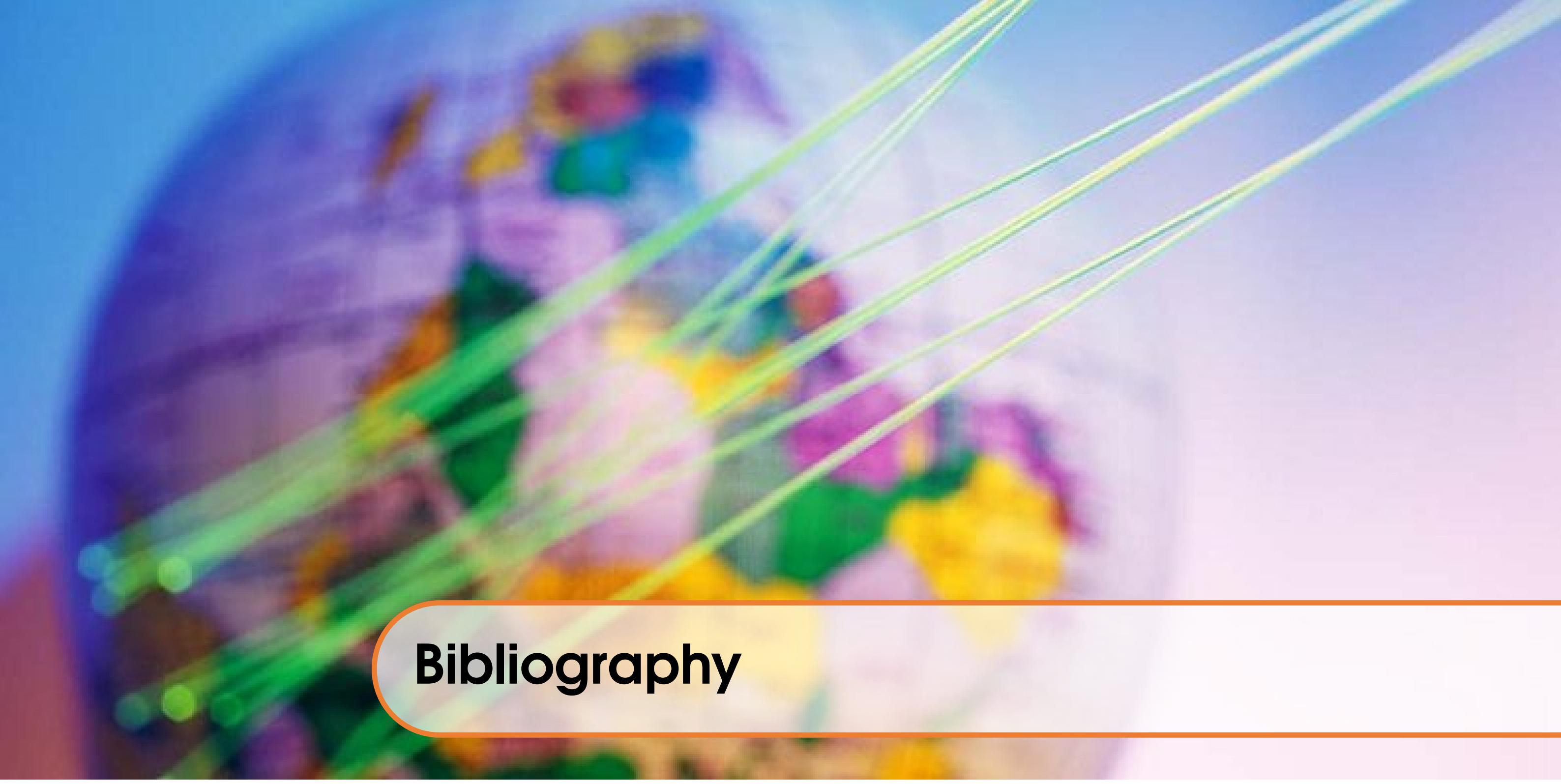
Liang et al. (2018) view China’s SCS as part of a broader effort to construct a “data-driven society.” According to Liang et al. (2018), China is embracing an approach to surveillance that goes beyond traditional surveillance efforts. They suggest the SCS embodies a “collect everything” approach to surveillance, in which automated methods are used to “track everything about everyone at all times” (Andrejevic & Gates, 2014, p. 190). Such an approach, Liang et al. suggest, eschews the use of sampling techniques in favour of large-scale, population-wide surveillance. They also note that traditional surveillance techniques are deployed for a specific purpose, whereas the collect everything approach is much less goal-directed. This is not to say that the collect everything approach is devoid of goals. Rather, what Liang et al. (2018) are suggesting is that the scope and scale of China's data collection efforts allows goals to arise as a result of surveillance. This contrasts with the more traditional approach where surveillance efforts are driven by predefined goals.

What is important here is the idea that data collection efforts can be characterized in such a way as to support the distinction between different kinds of surveillance. Liang et al. (2018) are interested in the distinction between traditional forms of surveillance and those associated with the SCS; however, other kinds of distinction may also be important (e.g., the distinction between commercial and government surveillance). One area for future research is thus to develop a better understanding of the 'design space' for data collection efforts. In particular, it would be useful to understand what types of data collection effort are feasible for different countries, in the sense of being compatible with nation-specific normative and legislative constraints. It would also be useful to understand how different kinds of data collection effort influence the nature of AI-oriented research. Does, for example, a collect everything type approach yield more opportunities for AI innovation, and does its eschewal of sampling techniques help or hinder the attempt to model population parameters? Clearly, the collect everything type approach referred to by Liang et al. is unlikely to be acceptable to liberal democracies. This, however, directs attention to the features of more limited data collection efforts. If, for example, sampling techniques are a feature of the data collection effort, then it is important that we understand how such techniques can be used to maximum effect. This speaks to some of the issues that were raised in Section 5.1 in respect of the notion of data value. In particular, we need to understand how data collection efforts can be optimized for specific purposes while respecting a variety of constraints. If the goal is to minimize the data collection effort so as to protect individual privacy, then it would be useful to know how much data we need to collect in order to deliver a good-enough AI solution. In parallel with the conventional notion of statistical power, it would help to have methods of determining the 'power' of a given data collection technique relative to the desired performance profile of an AI solution.

In Section 2, we saw how the vision of a single Open Internet, formulated along libertarian lines, is being challenged by the emergence of alternative Internet models. Such models threaten to overturn the traditional image of the online environment as a universal information space that serves the interests of humanity in a more or less uniform fashion. In its place, we have the image of a fragmented Internet. Here, the online world is riven by ideological differences and global power struggles—the very same tectonic forces that sculpt the topography of the real-world geopolitical map. In future work, it will be important to understand how different Internet models arise and how they are likely to evolve across time. Government regulation is no doubt important in facilitating the emergence of different Internet models and thus driving the process of Internet 'speciation', but such regulatory efforts are likely to be informed by pre-existing features of the geopolitical map (e.g., different countries may have different views about the problems and opportunities associated with the Internet). In addition, it is far from clear that similar regulatory interventions will have the same effect in different geopolitical constituencies (see Section 5.6). There are also a number of complex feedback loops to consider. Different Internet models may arise as a result of regulatory differences, but such models may also call for tailored regulatory responses, thereby accentuating the differences between regulatory regimes and the Internet models they support. In some cases, such feedback loops may hinder efforts to isolate cause–effect relationships. For example: Is the Internet changing in response to policies, or are the policies changing in response to the Internet?

Additional complexity can be found in the attempt to analyse particular areas of government policy. In this report, we made a rough distinction between four policy areas, namely, cyber, AI, social, and data governance. Moving forward, however, it is unclear whether any strict distinction between these governance areas can be sustained. In China, for example, it is increasingly hard to maintain a strict separation between social governance and cyber governance. In fact, inasmuch as the Internet has become an intrinsic part of social life in contemporary societies, it is increasingly hard to distinguish where the realm of the social stops and the realm of the cyber begins. Similarly, we have seen how issues of social governance—in the form of the SCS—are likely to have knock-on effects in terms of the availability of personal data and the development of AI capabilities. Such forms of interdependence highlight the importance of interdisciplinary collaboration. To fully understand the fragmentation of the Internet (and perhaps forestall the impending apocalypse), we need to embrace ideas and methods that cross-cut the traditional divisions between academic disciplines. To fully understand one kind of fragmentation, a new kind of fusion must be found.

This page was intentionally left blank.



Bibliography

Adelmann, P. K., & Zajonc, R. B. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology*, *40*(1), 249–280.

Allen, G. C. (2019). *Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security*. Center for a New American Security. Washington D.C., USA. Retrieved from https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance & Society*, *12*(2), 185–196.

Arya, S., Eckel, C., & Wichman, C. (2013). Anatomy of the credit score. *Journal of Economic Behavior & Organization*, *95*, 175–185.

Avena-Koenigsberger, A., Goñi, J., Solé, R., & Sporns, O. (2015). Network morphospace. *Journal of The Royal Society Interface*, *12*(103), 20140881.

Bailey, K. D. (1994). *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks, California, USA: Sage.

Barton, D., Woetzel, J., Seong, J., & Tian, Q. (2017). *Artificial Intelligence: Implications for China*. Mckinsey Global Institute. Retrieved from https://www.mckinsey.com/featured-insights/china/artificial-intelligence-implications-for-china

Baumeister, R. F., & Leary, M. R. (1995). The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological Bulletin*, *117*(3), 497–529.

Beijing Innovation Center for Future Chips. (2018). *White Paper on AI Chip Technologies*. Tsinghua University. Beijing, China. Retrieved from https://www.080910t.com/downloads/AI%20Chip%202018%20EN.pdf

Buckley, C., & Mozur, P. (2019). How China Uses High-Tech Surveillance to Subdue Minorities. The New York Times. Retrieved from https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html

Cadell, C. (2018). Beijing to build $2 billion AI research park: Xinhua. Retrieved from https://uk.reuters.com/article/us-china-artificial-intelligence/beijing-to-build-2-billion-ai-research-park-xinhua-idUKKBN1ES0B8

Chen, S. (2017). The world's next fastest supercomputer will help boost China's growing sea power. Retrieved from https://www.scmp.com/news/china/society/article/2107796/worlds-next-fastest-supercomputer-will-help-boost-chinas-growing

Chesney, R., & Citron, D. K. (forthcoming). Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review*. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954.

China Electronics Standardization Institute. (2018). *White Paper on Artificial Intelligence Standardization*. English translation: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-excerpts-chinas-white-paper-artificial-intelligence-standardization/.

China Institute for Science and Technology Policy. (2018). *China AI Development Report*. Tsinghua University. Beijing, China. Retrieved from http://www.sppm.tsinghua.edu.cn/eWebEditor/UploadFile/China_AI_development_report_2018.pdf

Chorzempa, M., Triolo, P., & Sacks, S. (2018). *China's Social Credit System: A Mark of Progress or a Threat to Privacy?* Peterson Institute for International Economics. Washington D.C., USA. Retrieved from https://ideas.repec.org/p/iie/pbrief/pb18-14.html

Chu, M.-c. M. (2013). *The East Asian Computer Chip War*. Abingdon, Oxon, UK: Routledge.

Chu, M.-c. M. (2016). China's ambitions in the semiconductor sphere and Taiwan's dilemma. Retrieved from https://www.bbc.com/zhongwen/trad/china/2016/05/160506_china_semicounductor_business

Clark, A. (2013). Expecting the world: Perception, prediction, and the origins of human knowledge. *The Journal of Philosophy*, *110*(9), 469–496.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action and the Embodied Mind*. New York, New York, USA: Oxford University Press.

Clinton, H. R. (2010). Remarks on Internet Freedom. Retrieved from https://2009-2017.state.gov/secretary/20092013clinton/rm/2010/01/135519.htm

Condon, S. (2017). TPU is 15x to 30x faster than GPUs and CPUs, Google says. SDNet. Retrieved from https://www.zdnet.com/article/tpu-is-15x-to-30x-faster-than-gpus-and-cpus-google-says/

Cook, S. (2018). China's Cyber Superpower Strategy: Implementation, Internet Freedom Implications, and U.S. Responses. Retrieved from https://freedomhouse.org/article/china-s-cyber-superpower-strategy-implementation-internet-freedom-implications-and-us

Cooper, J. (2007). *Cognitive Dissonance: Fifty Years of a Classic Theory*. London, UK: Sage Publications Ltd.

Cotterell, C. (2016). How AI Is Shaking Up the Chip Market. Retrieved from https://www.wired.com/2016/10/ai-changing-market-computer-chips/

Craver, C. (2007). Constitutive explanatory relevance. *Journal of Philosophical Research*, *32*, 3–20.

Creemers, R. (2017). Cyber China: Upgrading Propaganda, Public Opinion Work and Social Management for the Twenty-First Century. *Journal of Contemporary China*, *26*(103), 85–100.

Creemers, R. (2018a). China's Social Credit System: An Evolving Practice of Control. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175792.

Creemers, R. (2018b). Cyber-Leninism: The Political Culture of the Chinese Internet. In M. Price & N. Stremlau (Eds.), *Speech and Society in Turbulent Times: Freedom of Expression in Comparative Perspective* (pp. 255–273). Cambridge, UK: Cambridge University Press.

Cyberspace Administration of China. (2015). *Internet User Account Name Management Regulations*. English translation: https://chinacopyrightandmedia.wordpress.com/2015/02/04/internet-user-account-name-management-regulations/.

Dai, X. (2018). Toward a reputation state: The social credit system project of China. Available at SSRN: https://ssrn.com/abstract=3193577.

Deloitte. (2017). *Hitting the accelerator: the next generation of machine-learning chips*. Deloitte Touche Tohmatsu Limited. London, UK. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/global/Images/infographics/technologymediatelecommunications/gx-deloitte-tmt-2018-nextgen-machine-learning-report.pdf

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, Masachussetts, USA: MIT Press.

Ding, J. (2018). *Deciphering China's AI Dream*. Future of Humanity Institute, University of Oxford. Oxford, UK. Retrieved from https://www.fhi.ox.ac.uk/deciphering-chinas-ai-dream/

Donnan, S. (2017). Trump Blocks US Chipmaker's Sale to China-backed Buyer. Financial Times. Retrieved from https://www.ft.com/content/d2924226-98ce-11e7-a652-cde3f882dd7b

Dutton, T. (2018). An Overview of National AI Strategies. Retrieved from https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

European Union Chamber of Commerce in China, & Sinolytics. (2019). *The Digital Hand: How China's Corporate Social Credit System Conditions Market Actors*. European Union Chamber of Commerce in China. Retrieved from https://www.europeanchamber.com.cn/en/press-releases/3045/european_chamber_report_on_china_s_corporate_social_credit_system_a_wake_up_call_for_european_business_in_china

Fabre, G. (2018). *China's digital transformation. Why is artificial intelligence a priority for chinese R&D?* Fondation Maison des sciences de l'homme. Paris, France. Retrieved from https://halshs.archives-ouvertes.fr/halshs-01818508v2/document

Feng, C. (2018). AI-Chip Unicorn Wins $2.5 Billion Valuation. Retrieved from https://www.caixinglobal.com/2018-06-20/ai-chip-unicorn-wins-25-billion-valuation-101273597.html

Feola, J. (2018). China Exports Facial Scan Tech to Zimbabwe, Launches First "AI Technology Entry to Africa". Retrieved from https://radiichina.com/china-exports-facial-scan-tech-to-zimbabwe-launches-first-ai-technology-entry-to-africa/

Fischer, S.-C. (2018). *Artificial Intelligence: China's High-Tech Ambitions*. Center for Security Studies, ETH Zürich. Zürich, Switzerland. Retrieved from https://ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSSAnalyse220-EN.pdf

Floridi, L. (2014). *The 4th Revolution: How the Infosphere is Reshaping Human Reality*. Oxford, UK: Oxford University Press.

Gentle, M. (2018). China's data-privacy law vs. GDPR. Retrieved from https://medium.com/the-balance-of-privacy/chinas-data-privacy-law-vs-gdpr-566fde8c213c

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, Massachusetts, USA: MIT Press.

Government of the United Kingdom. (2019). *Online Harms White Paper*. London, UK: Her Majesty's Stationery Office.

Grenoble, R. (2017). Welcome To The Surveillance State: China's AI Cameras See All. Retrieved from https://www.huffingtonpost.com.au/entry/china-surveillance-camera-big-brother_n_5a2ff4dfe4b01598ac484acc?guccounter=1

Hall, W., & Pesenti, J. (2017). *Growing the Artificial Intelligence Industry in the UK*. HM Majesty's Government. London, UK. Retrieved from https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk

Hatton, C. (2015). China 'social credit': Beijing sets up huge system. BBC News. Retrieved from https://www.bbc.co.uk/news/world-asia-china-34592186

Hedström, P., & Bearman, P. (Eds.). (2009). *The Oxford Handbook of Analytical Sociology*. Oxford, UK: Oxford University Press.

Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, *36*, 49–67.

Hendler, J., & Berners-Lee, T. (2010). From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. *Artificial Intelligence*, *174*, 156–161.

Henley, J. (2019). Ten cities ask EU for help to fight Airbnb expansion. The Guardian. Retrieved from https://www.theguardian.com/cities/2019/jun/20/ten-cities-ask-eu-for-help-to-fight-airbnb-expansion

Hinsliff, G. (2018). Airbnb and the so-called sharing economy is hollowing out our cities. The Guardian. Retrieved from https://www.theguardian.com/commentisfree/2018/aug/31/airbnb-sharing-economy-cities-barcelona-inequality-locals

Hinton, G. E. (2010). Learning to represent visual input. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1537), 177–184.

Hiranand, R. (2018). Shenzhen's jaywalkers get scolded on WeChat and shamed on public screens. Retrieved from https://www.abacusnews.com/future-tech/shenzhens-jaywalkers-get-scolded-wechat-and-shamed-public-screens/article/2138928

HM Treasury. (2018). *The economic value of data: discussion paper*. HM Treasury. London, UK. Retrieved from https://www.gov.uk/government/publications/the-economic-value-of-data-discussion-paper

Hoffman, S. (2017). *Programming China: The Communist Party's autonomic approach to managing state security*. Mercator Institute for China Studies. Berlin, Germany. Retrieved from https://www.merics.org/sites/default/files/2017-12/171212_China_Monitor_44_Programming_China_EN__0.pdf

Hong, Y. (2018). Responses and explanations to the five major concerns of the Personal Information Security Code. Retrieved from https://mp.weixin.qq.com/s/rSW-Ayu6zNXw87itYHcPYA

Huawei. (2018). Video Surveillance as the Foundation of "Safe City" in Kenya. Retrieved from https://www.huawei.com/uk/industry-insights/technology/digital-transformation/video/video-surveillance-as-the-foundation-of-safe-city-in-Kenya

Hurley, M., & Adebayo, J. (2016). Credit scoring in the era of big data. *Yale Journal of Law and Technology*, *18*(1), 148–216.

Jing, M., & Dai, S. (2017). China recruits Baidu, Alibaba and Tencent to AI 'national team'. Retrieved from https://www.scmp.com/tech/china-tech/article/2120913/china-recruits-baidu-alibaba-and-tencent-ai-national-team

Johnson-Roberson, M., Barto, C., Mehta, R., Sridhar, S. N., Rosaen, K., & Vasudevan, R. (2017). Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? In *IEEE International Conference on Robotics and Automation*, Singapore: IEEE.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343.

Knockel, J., Ruan, L., Crete-Nishihata, M., & Deibert, R. (2018). *(Can't) Picture This: An Analysis of Image Filtering on WeChat Moments*. The Citizen Lab, University of Toronto. Toronto, Canada. Retrieved from https://citizenlab.ca/2018/08/cant-picture-this-an-analysis-of-image-filtering-on-wechat-moments/

Kostka, G. (2019). China's social credit systems and public opinion: Explaining high levels of approval. *New Media & Society*, *21*(7), 1565–1593.

Kovel, J. (2007). *The Enemy of Nature: The End of Capitalism or the End of the World?* London, UK: Zed Books Ltd.

Landes, D. (2000). *Revolution in Time: Clocks and the Making of the Modern World*. London, UK: Viking Press.

Larson, C. (2018). China's massive investment in artificial intelligence has an insidious downside. Retrieved from https://www.sciencemag.org/news/2018/02/china-s-massive-investment-artificial-intelligence-has-insidious-downside

Lauer, J. (2017). *Creditworthy: A History of Consumer Surveillance and Financial Identity in America*. New York, New York, USA: Columbia University Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lew, L. (2017). Fast and furious: Chinese unicorns to overtake American counterparts says BCG report. Retrieved from https://technode.com/2017/09/18/fast-and-furious-chinese-unicorns-to-overtake-american-counterparts-says-bcg-report/

Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a Data-Driven Society: China's Social Credit System as a State Surveillance Infrastructure. *Policy & Internet*, *10*(4), 415–453.

Liao, S.-K., Cai, W.-Q., Handsteiner, J., Liu, B., Yin, J., Zhang, L., . . . Liu, W.-Y. (2018). Satellite-relayed intercontinental quantum network. *Physical Review Letters*, *120*(3), 030501.

Lin, J., & Singer, P. W. (2017). China is opening a new quantum research supercenter. Retrieved from https://www.popsci.com/chinas-launches-new-quantum-research-supercenter/

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition. *Frontiers in Robotics and AI*, *5*(21), 1–22.

Lu, Y., Qian, D., Fu, H., & Chen, W. (2018). Will Supercomputers Be Super-Data and Super-AI Machines? *Communications of the ACM*, *61*(11), 82–87.

Lucas, L. (2019). Cold water hits China's AI industry. Financial Times. Retrieved from https://www.ft.com/content/973bfc08-a15f-11e9-a282-2df48f366f7d

Luo, Y. (2019). China Releases Draft Amendments to the Personal Information Protection Standard. Retrieved from https://www.insideprivacy.com/international/china/china-releases-draft-amendments-to-the-personal-information-protection-standard/

Madrigal, A. C. (2017). Inside Waymo's Secret World for Training Self-Driving Cars. The Atlantic. Retrieved from https://www.theatlantic.com/technology/archive/2017/08/inside-waymos-secret-testing-and-simulation-facilities/537648/

Maedche, A., & Staab, S. (2002). Measuring Similarity between Ontologies. In A. Gómez-Pérez & V. R. Benjamins (Eds.), (pp. 251–263). Sigüenza, Spain: Springer.

Marchionni, C. (2018). Mechanisms in economics. In S. Glennan & P. M. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 423–434). New York, New York, USA: Routledge.

Meissner, M. (2017). *China's Social Credit System: A big-data enabled approach to market regulation with broad implications for doing business in China*. Mercator Institute for China Studies. Berlin, German. Retrieved from https://www.merics.org/sites/default/files/2017-09/China%20Monitor_39_SOCS_EN.pdf

Meyer, E. T., & Schroeder, R. (2015). *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Cambridge, Massachusetts, USA: MIT Press.

Mistreanu, S. (2018). Life Inside China's Social Credit Laboratory. Retrieved from https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/

Mitteroecker, P., & Huttegger, S. M. (2009). The concept of morphospaces in evolutionary and developmental biology: Mathematics and metaphors. *Biological Theory*, *4*(1), 54–67.

Moore-Colyer, R. (2019). Samsung looks set to have 5nm chips ready for 2020. Retrieved from https://www.theinquirer.net/inquirer/news/3078674/samsung-5nm-chips-2020

Moore, S. K. (2019). Another Step Toward the End of Moore's Law. *IEEE Spectrum*, *56*(6), 9–10.

Moore, S. M. (2018). Apple, Huawei Both Claim First 7-nm Smartphone Chips. Retrieved from https://spectrum.ieee.org/nanoclast/semiconductors/processors/apple-huaweii-both-claim-first-7nm-smartphone-chips

Moreau, L. (2010). The foundations for provenance on the Web. *Foundations and Trends in Web Science*, *2*(2–3), 99–241.

Mozur, P. (2017). Beijing Wants A.I. to Be Made in China by 2030. *The New York Times*. Retrieved from https://www.nytimes.com/2017/07/20/business/china-artificial-intelligence.html

National Development and Reform Commission. (2017). Notice of the list of the first batch of demonstration cities for social credit system construction. Retrieved from http://www.ndrc.gov.cn/zcfb/zcfbtz/201801/t20180109_873397.html

National Science and Technology Council. (2016). *The National Artificial Intelligence Research and Development Strategic Plan*. Executive Office of the President of the United States. Washington D.C., USA. Retrieved from https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf

O'Hara, K., & Hall, W. (2018). *Four Internets: The Geopolitics of Digital Governance*. The Centre for International Governance Innovation. Waterloo, Ontario, Canada. Retrieved from https://www.cigionline.org/publications/four-internets-geopolitics-digital-governance

Odling-Smee, J., Laland, K. N., & Feldman, M. W. (2003). *Niche Construction: The Neglected Process in Evolution*. Princeton, New Jersey, USA: Princeton University Press.

Ohlberg, M., Ahmed, S., & Lang, B. (2017). *Central planning, local Experiments: The complex implementation of China's Social Credit System*. Mercator Institute for China Studies. Berlin, German. Retrieved from https://www.merics.org/sites/default/files/2018-03/171212_China_Monitor_43_Social_Credit_System_Implementation_1.pdf

Ojeda, E. O. (2019). *Gamifying trust: Can you win at life? China's Social Credit System and the rating revolution*. Lee Kuan Yew School of Public Policy, National University of Singapore. Singapore. Retrieved from https://scholarbank.nus.edu.sg/handle/10635/154274?mode=full

Pentland, A. (2014). *Social Physics: How Good Ideas Spread—The Lessons from a New Science*. New York, New York, USA: Penguin Press.

Persily, N. (2017). The 2016 US Election: Can democracy survive the Internet? *Journal of Democracy*, *28*(2), 63–76.

President's Council of Advisors on Science and Technology. (2017). *Ensuring Long-Term U.S. Leadership in Semiconductors*. Executive Office of the President of the United States. Washington D.C., USA. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_ensuring_long-term_us_leadership_in_semiconductors.pdf

Putin, V. (2017). 'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day. RT. Retrieved from https://www.rt.com/news/401731-ai-rule-world-putin/

PwC. (2016). *China's Impact on the Semiconductor Industry: 2015 Update*. PwC. Retrieved from https://www.pwc.com/gx/en/technology/pdf/china-semicon-2015-report-1-5.pdf

Qi, Y., & Xiao, J. (2018). Fintech: AI Powers Financial Services to Improve People's Lives. *Communications of the ACM*, *61*(11), 65–69.

Ridley, M. (2003). *The Red Queen: Sex and the Evolution of Human Nature*. New York, New York, USA: Perennial.

Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In E.-P. Lim & M. Winslett (Eds.), *Proceedings of the 2017 ACM Conference on Information and Knowledge Management* (pp. 797–806). Singapore: ACM.

Sacks, S. (2018a). China's Emerging Data Privacy System and GDPR. Retrieved from https://www.csis.org/analysis/chinas-emerging-data-privacy-system-and-gdpr

Sacks, S. (2018b). New China Data Privacy Standard Looks More Far-Reaching than GDPR. Retrieved from https://www.csis.org/analysis/new-china-data-privacy-standard-looks-more-far-reaching-gdpr

Sacks, S., Triolo, P., & Webster, G. (2017). Beyond the Worst-Case Assumptions on China's Cybersecurity Law. Retrieved from https://www.newamerica.org/cybersecurity-initiative/blog/beyond-worst-case-assumptions-chinas-cybersecurity-law/

Sanger, D. E., Barboza, D., & Perlroth, N. (2013). Chinese Army Unit Is Seen as Tied to Hacking Against U.S. The New York Times. Retrieved from https://www.nytimes.com/2013/02/19/technology/chinas-army-is-seen-as-tied-to-hacking-against-us.html

Schmitz, R. (2017). What's Your 'Public Credit Score'? The Shanghai Government Can Tell You. Retrieved from https://www.npr.org/sections/parallels/2017/01/03/507983933/whats-your-public-credit-score-the-shanghai-government-can-tell-you?t=1567086549587

Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N. R., Van de Velde, W., & Weilinga, B. (2000). *Knowledge Engineering and Management: The CommonKADS Methodology*. Cambridge, Massachusetts, USA: MIT Press.

Segal, A. (2018). When China Rules the Web: Technology in Service of the State. *Foreign Affairs*, *97*(5), 10–18.

Shen, C. F. (2019). Social Credit System in China. In C. Echle, K. Naumann, & M. Sarmah (Eds.), *Digital Asia* (pp. 21–31). Singapore: Konrad-Adenauer-Stiftung Ltd.

Shen, H. (2018). Building a Digital Silk Road? Situating the Internet in China's Belt and Road Initiative. *International Journal of Communication*, *12*, 2683–2701.

Shi, M., Sacks, S., Chen, Q., & Webster, G. (2019). Translation: China's Personal Information Security Specification. Retrieved from https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinas-personal-information-security-specification/

Shotwell, P. (2008). *The Game of Go: Speculations on its Origins and Symbolism in Ancient China*. American Go Association. New York, New York, USA. Retrieved from https://www.usgo.org/sites/default/files/bh_library/originsofgo.pdf

Sieck, W. R., Rasmussen, L., & Smart, P. R. (2010). Cultural Network Analysis: A Cognitive Approach to Cultural Modeling. In D. Verma (Ed.), *Network Science for Military Coalition Operations: Information Extraction and Interaction* (pp. 237–255). Hershey, Pennsylvania, USA.: IGI Global.

Silver, D., & Hassabis, D. (2016). AlphaGo: Mastering the ancient game of Go with Machine Learning. Retrieved from https://ai.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., . . . Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359.

Síthigh, D. M., & Siems, M. (2019). *The Chinese Social Credit System: A Model for other Countries?* European University Institute. Fiesole, Italy. Retrieved from https://cadmus.eui.eu/handle/1814/60424

Smart, P. R. (2018a). Knowledge Machines. *The Knowledge Engineering Review*, *33*(e11), 1–26.

Smart, P. R. (2018b). Mandevillian intelligence. *Synthese*, *195*(9), 4169–4200.

Smart, P. R. (2020). Artificial Economics. In T. Shanahan & P. R. Smart (Eds.), *Blade Runner 2049: A Philosophical Exploration* (pp. 185–205). Abingdon, Oxon, UK: Routledge.

Smart, P. R., Madaan, A., & Hall, W. (2019). Where the Smart Things Are: Social Machines and the Internet of Things. *Phenomenology and the Cognitive Sciences*, *18*(3), 551–575.

Smart, P. R., & Shadbolt, N. R. (2014). Social Machines. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (pp. 6855–6862). Hershey, Pennsylvania, USA: IGI Global.

Smart, P. R., Sieck, W. R., Braines, D., Huynh, T. D., Sycara, K., & Shadbolt, N. R. (2010). Modelling the Dynamics of Collective Cognition: A Network-Based Approach to Socially-Mediated Cognitive Change. In *4th Annual Conference of the International Technology Alliance (ACITA'10)*, London, UK.

Smart, P. R., Simperl, E., & Shadbolt, N. R. (2014). A Taxonomic Framework for Social Machines. In D. Miorandi, V. Maltese, M. Rovatsos, A. Nijholt, & J. Stewart (Eds.), *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society* (pp. 51–85). Berlin, Germany: Springer.

Standing Committee of the National People's Congress. (2017). *Cybersecurity Law of the People's Republic of China*. English Translation: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-cybersecurity-law-peoples-republic-china/.

State Council. (2007). *Certain Opinions from the General Office of the State Council on Construction of the Social Credit System*. English Translation: https://chinacopyrightandmedia.wordpress.com/2007/03/23/state-council-general-office-some-opinions-concerning-the-construction-of-a-social-credit-system/.

State Council. (2014). *Planning Outline for the Construction of a Social Credit System (2014-2020)*. English translation: https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/.

State Council. (2015). *Made in China 2025*. English translation: http://www.cittadellascienza.it/cina/wp-content/uploads/2017/02/IoT-ONE-Made-in-China-2025.pdf.

State Council. (2016a). *Distribution Notice from the State Council on the Fifteenth Five Year National Informatization Plan*. Retrieved from http://www.gov.cn/zhengce/content/2016-12/27/content_5153411.htm

State Council. (2016b). *Guiding Opinions concerning Establishing and Perfecting Incentives for Promise-keeping and Joint Punishment Systems for Trust-Breaking, and Accelerating the Construction of Social Sincerity*. English Translation: https://chinacopyrightandmedia.wordpress.com/2016/05/30/state-council-guiding-opinions-concerning-establishing-and-perfecting-incentives-for-promise-keeping-and-joint-punishment-systems-for-trust-breaking-and-accelerating-the-construction-of-social-sincer/.

State Council. (2017). *New Generation Artificial Intelligence Development Plan*. English translation: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.

Tan, A. (2018). China to open $2.1bn AI tech park in Beijing. Retrieved from https://www.computerweekly.com/news/450432800/China-to-open-21bn-AI-tech-park-in-Beijing

Tao, H. (2017). Zhima Credit does not share user scores or data. Financial Times. Retrieved from https://www.ft.com/content/ec4a2a46-c577-11e7-a1d2-6786f39ef675

Tao, L. (2019). SenseNets: the facial recognition company that supplies China's Skynet surveillance system. South China Morning Post. Retrieved from https://www.scmp.com/tech/science-research/article/3005733/what-you-need-know-about-sensenets-facial-recognition-firm

The Economist. (2017a). China may match or beat America in AI. Retrieved from https://www.economist.com/business/2017/07/15/china-may-match-or-beat-america-in-ai

The Economist. (2017b). Is Margrethe Vestager championing consumers or her political career? The Economist. Retrieved from https://www.economist.com/business/2017/09/14/is-margrethe-vestager-championing-consumers-or-her-political-career

Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2016). Credit scoring with social network data. *Marketing Science*, *35*(2), 234–258.

Woetzel, J., Seong, J., Wang, K. W., Manyika, J., Chui, M., & Wong, W. (2017). *China's Digital Economy: A Leading Global Force*. McKinsey Global Institute. Retrieved from https://www.mckinsey.com/featured-insights/china/chinas-digital-economy-a-leading-global-force

Wong, K. L. X., & Dobson, A. S. (2019). We're just data: Exploring China's social credit system in relation to digital platform ratings cultures in Westernised democracies. *Global Media and China*, *4*(2), 220–232.

Xia, H., & Yang, H. (2018). Is Last-Mile Delivery a 'Killer App' for Self-Driving Vehicles? *Communications of the ACM*, *61*(11), 70–75.

Yang, Y., & Murgia, M. (2019). Data leak reveals China is tracking almost 2.6m people in Xinjiang. Financial Times. Retrieved from https://www.ft.com/content/9ed9362e-31f7-11e9-bb0c-42459962a812

Ylikoski, P. (2018). Social mechanisms. In S. Glennan & P. M. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 401–412). New York, New York, USA: Routledge.

Yu, X. (2016). Satellite Micius to help create £5.7bn quantum communication market. The Telegraph. Retrieved from https://www.telegraph.co.uk/sponsored/china-watch/technology/12212871/china-satellite-micius-gobi-desert.html

Zaagman, E. (2018). China's Computing Ambitions. *Communications of the ACM*, *61*(11), 40–41.

Zhu, J., Huang, T., Chen, W., & Gao, W. (2018). The Future of Artificial Intelligence in China. *Communications of the ACM*, *61*(11), 44–45.

This page was intentionally left blank.

Acronyms

AAAI	Association for the Advancement of Artificial Intelligence
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AIAC	Artificial Intelligence Accelerator Chip
CAC	Cyberspace Administration of China
CAICT	China Academy of Information and Communications Technology
CCP	Chinese Communist Party
CFIUS	Committee on Foreign Investment in the United States
CNKI	China National Knowledge Infrastructure
CPU	Central Processing Unit
CRC	Credit Reference Center
EU	European Union
eWTP	Electronic World Trade Platform
FPGA	Field Programmable Gate Array
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
ICT	Information and Communication Technology
MIIT	Ministry of Industry and Information Technology

MOST	Ministry of Science and Technology
NCISP	National Credit Information Sharing Platform
NDRC	National Development and Reform Commission
NECIPS	National Enterprise Credit Information Publicity System
PBoC	People's Bank of China
SCP	Social Credit Plan
SCS	Social Credit System
SMIC	Semiconductor Manufacturing International Corporation
SoC	System-On-Chip
TPU	Tensor Processing Unit
TSMC	Taiwan Semiconductor Manufacturing Company
UMC	United Microelectronics Corporation
US	United States

A — Appendix A: Research Areas

The following sections describe areas for future research that were identified as part of the effort to produce the present report. The research areas are grouped into the following categories:

- Cyber Governance and the Four Internets (Section A.1),
- Data Governance and Value (Section A.2),
- Cognitive Modelling and Cyber Epistemology (Section A.3),
- Understanding China (Section A.4), and
- Artificial Intelligence (Section A.5).

A.1 Cyber Governance and the Four Internets

A.1.1 The Cyber Governance Morphospace

Section 2 highlighted some of the features of the Internet models discussed by O'Hara and Hall (2018). Future work could seek to extend this 'featural characterization' of the four Internets by developing a taxonomic framework for Internet models. Such an approach relies on the use of classification techniques (see Bailey, 1994) to identify the dimensions along which different Internet models can be seen to vary. If successful, the result will be a multidimensional space in which all extant Internet models can be located. The values assigned to an Internet model with respect to these dimensions specifies its position within the multi-dimensional space, and this position serves as the basis for the computation of distance metrics that tell us how close (or similar) one Internet model is to another.

The development of a taxonomic framework for Internet models comes with an added bonus. In addition to providing the basis for comparative analyses between extant Internet models, the multidimensional characterization also allows us to explore the 'universe' of all possible Internet models. This is because each point within the multidimensional space identifies a particular kind of Internet model. Many such models will not, of course, exist. The universe of Internet models, as it is currently constituted, is sparsely populated by a mere four models. But the Internet is arguably at an early stage of its evolution, and we should not assume that the Internet's future is defined by the shape (or form) of models that exist at the time of writing.

The reference to evolution (and form) is important, for multidimensional spaces are a feature of work that seeks to chart the trajectory of evolutionary processes through a space of morphological possibilities. Indeed, the notion of a multidimensional space is formally equivalent to what is called a *morphospace* in the biological sciences (Mitteroecker & Huttegger, 2009). Historically, work in this area has been confined to the biological realm; however, recent work has sought to apply the morphospace concept to other areas (e.g., Avena-Koenigsberger, Goñi, Solé, & Sporns, 2015; Smart, Simperl, & Shadbolt, 2014). In relation to the present research effort, the aim is to apply the morphospace concept to the realm of cyber governance and Internet 'morphology'. In other words, the aim is to chart (and explore) the universe of Internet models based on a coordinate system that reflects ideologically-inflected views of what the Internet should be and whose interests it should serve.

Morphospaces represent a universe of forms. But morphospaces, themselves, come in different varieties. A crucial distinction is between morphospaces that represent the features of forms and those that represent the parameters of generative processes that give rise to those forms. Let us refer to the former as featural morphospaces and the latter as generative morphospaces. The foregoing discussion targets the notion of a featural morphospace, for the aim is to describe the features of Internet models (or Internet morphologies). A consideration of generative morphospaces is, however, also likely to be important. The reason for this is that it is unclear whether different Internet models should be characterised in terms of the features of Internet-related processes or whether the models should be characterised relative to the way in which these features emerge.

To help us understand this, let us direct our attention to the properties of a particular Internet model, namely, the Paternal Internet, which is emblematic of China's approach to cyber governance. As we have seen, China's approach to cyber governance is characterized by greater government control over the Internet ecosystem (see Section 2.4). This conflicts with the foundational principles of the Internet in liberal, market-based democracies. Relative to Western models, then, we see a greater level of 'top-down' control when it comes to various aspects of the online data ecology, such as what information appears online and who has access to particular bodies of data. This contrasts with a more Western approach, in which issues of access and censorship are resolved from the 'bottom-up' (e.g., via market forces and market mechanisms).

Let us now imagine a state-of-affairs in which the Commercial Internet is the only game in town. That is to say, imagine there are no geopolitical differences in the approach to cyber governance—all approaches converge on that associated with the Commercial Internet. Also suppose that US Internet companies are driven to engage in censorship and information manipulation as a response to market forces and that they do so to the same extent as that observed in the case of (what we would otherwise call) the Paternal Internet. Does this mean that the Commercial Internet has morphed into the Paternal Internet? Or do we still have a Commercial Internet that merely resembles the Paternal Internet? In support of the latter proposal, we might appeal to the *way* in which certain features (e.g., censorship) arise, as opposed to the mere existence of those features. We might thus insist that we still have a Commercial Internet on the grounds that, relative to the present example, censorship arises as a result of market forces rather than any attempt at top-down government control. Perhaps, then, what is important for classificatory purposes is not so much the precise location of a given Internet model within the aforementioned morphospace; rather, what matters is how it arrived at that location. Did it come about as a result of market mechanisms, or was it the result of an explicit attempt at government control? Here we see the importance of the featural/generative distinction mentioned above. For a featural morphospace that focuses on censorship will not distinguish between the Commercial and Paternal Internets; a generative morphospace will.

A.1.2 Internet Evolution and the Geopolitical Niche

From an evolutionary perspective, the Open Internet model is arguably *open* to 'invasion' by competing strategies, just as a population of pure 'doves' is prone to invasion by birds of a more 'hawkish' variety. Does this signal the ultimate demise of the Open Internet, or is there an evolutionarily stable strategy that arises as a result of context-dependent shifts in the 'fitness' of specific Internet models? Is there a Nash equilibrium for global cyber governance and, and if so, what would it look like?

The appeal to issues of fitness (e.g., the fitness of an Internet model), helps to focus attention on the dynamics of Internet evolution. Inasmuch as we conceive of Internet models as particular strategies, then we can begin to see how the success of different models (or strategies) varies as a function of the popularity of other models. A single 'hawk' will fare very well in a population of pure 'doves'. However, as the population comes to be dominated by hawks, the benefits associated with the hawkish strategy start to diminish and costs start to become the predominant concern.

When it comes to issues of Internet evolution, it pays to bear in mind the actions taken by nation states to manipulate the structure of the geopolitical landscape, and thereby alter the relative fitness of particular approaches to cyber governance. To help us understand this consider that in biology an organism's fitness is a relational property, reflecting the 'match' between a particular phenotype and a particular environment. In an effort to maximize fitness, organisms can adapt their phenotype to suit the environment—that is to say, they adapt or evolve in response to evolutionary pressures. But organisms do not simply evolve in response to their environments, they also *intervene* in the environment so as to shape the course of their species' evolutionary trajectory. In this sense, evolution is an active process. To increase one's fitness, one can either change oneself to suit the world, or one can change the world to suit oneself. This latter strategy is what is commonly referred to as *niche construction* (Odling-Smee, Laland, & Feldman, 2003).[15]

This raises a number of important issues. To what extent, for example, do the ideological proponents of different Internet models seek to reshape the prevailing geopolitical landscape in ways that alter the fitness of their preferred model and thus its prospects for survival? Are some nation states more active in doing this? And, if so, what impact does it have on the likely long-term success of certain models?

In addressing these issues, China's Belt and Road Initiative is likely to be an important focus of attention. By contributing to the development of third world countries, China facilitates the emergence of new markets for its own products. In addition, the Belt and Road initiative provides important opportunities for China to export its approach to cyber governance. From a niche construction perspective, such efforts emerge as a form of ideological 'terraforming'—they are, in essence, an attempt to till the geopolitical terrain so as to improve the fecundity of a particular approach to cyber governance.

A.1.3 Cyber Ecological Engineering: Shaping the (Techno) Social

In Section 3.4.6, we encountered the idea that China's approach to cyber governance helps to shape the Internet in a manner that is consistent with an authoritarian agenda. In particular, we suggested that China's Internet is dominated by a limited number of home-grown (i.e., Chinese) commercial giants (e.g., Baidu, Alibaba, Tencent) and that such a state-of-affairs has arisen (in

[15] This distinction between passive (adaptation) and active (niche construction) modes in evolutionary theory parallels the distinction between perception and action in free energy formulations of brain function. The significance of this will become clearer in Section A.5.3. For present purposes, however, we should note that, according to the free energy approach, perception involves changing a (cognitive) model to reflect sensory input, whereas action involves changing the sensory environment to match a (cognitive) model. In both cases, the 'fitness' of a model is determined by the match between model parameters and the nature of sensory reality. This is represented by free energy, which is at a minimum when the match is optimal. The basic computational imperative of the biological brain is then to minimize free energy, which it does either by changing itself to match the world (perception) or changing the world to match itself (action) (see Linson, Clark, Ramamoorthy, & Friston, 2018, for a more detailed discussion).

part) as a result of China's protectionist policies. We also suggested that a commercial ecosystem dominated by a limited number of companies is favourable to the Chinese government. This is because (1) the companies are Chinese (rather than American), and (2) it is easier to regulate three or four large companies than it is lots of small ones.

From a niche construction perspective, then, China has adopted policies that have helped to shape its local Internet ecology in a manner that befits its own ideologically-inflected approach to cyber governance. This is what we might call *cyber ecological engineering*.

Another example of cyber ecological engineering was mentioned in Section 2.4 in the form of the so-called "retweet rule." As we saw, the retweet rule is intended to prevent the spread of 'fake' news—if false or misleading information is propagated (retweeted) more than 500 times, then the originator is subject to custodial penalties. As noted in Section 2.4, the introduction of this rule increases the risk associated with certain kinds of online social media systems. Open systems, where individual posts can be accessed by anyone and propagated indiscriminately, are particularly hazardous. The result is a potential shift in the social media landscape, from open systems, like Weibo, to more private messaging systems, like WeChat (see Creemers, 2017). The latter systems retard the rate of information dissemination, thereby limiting the spread of 'harmful' information. They also, as an added bonus, give the Chinese leadership time to evaluate and respond to subversive or politically-sensitive information.

Such examples help us understand how the CCP shapes the nature of the online ecology in a way that ensures its continued survival.[16] They also help us appreciate the bidirectional influences between governmental processes and the Internet—the Internet informs government policy, while government policy influences the nature of the online ecology. The development of these ideas and the identification of additional forms of cyber ecological engineering (both inside and outside China) is an important topic for future research.

A.1.4 Transcending the Geopolitical Landscape: The Space Internet

In Section 2, we saw how the erstwhile dominance of the Open Internet is being challenged by the emergence of more recent Internet models, namely, the Commercial, Bourgeois, and Paternal Internet models. It is, as yet, unclear how the 'battle' between these various Internet models will play out. In particular, we do not know if one model will triumph over the others to become the new 'king of the models'. Somewhat worryingly for proponents of the Open Internet, the Paternal Internet looks set to increase its influence. Not only is it the model adopted by the world's most populous nation, but it is also at risk of being exported to other countries as a result of China's Belt and Road Initiative. What is worse, the Open Internet is under threat from the Commercial Internet, which challenges the authority of the Open Internet in its own geopolitical backyard (see Section 2.3).

A characteristic feature of the Open Internet is its preference for technological solutions to problems. Proponents of the Open Internet are not oblivious to some of the issues raised by the Internet—they recognize, for example, the genuine nature of the threat posed by fake news. However, technological innovation tends to be the preferred solution strategy for proponents of the Open Internet. This does not mean that proponents of the Open Internet are anti-regulation; it is simply that technological innovation (rather than regulatory fiat) is the preferred approach to tackling problems. Indeed, to some extent, it might be thought that such problems are one of the major drivers *for* technological innovation. That is to say, technological innovation occurs as a result of efforts to address specific problems, but problems also arise as the result of technological innovation. The problem of fake news, for example, arises (or is at least exacerbated) by the

[16] Remember: The fitness of any system is determined by the match between itself and the environment in which it finds itself. Such matches can be improved by changing oneself to suit the environment (adaptation) or by changing the environment to suit oneself (niche construction).

invention of the Internet. The result is an ongoing cycle of technological innovation, in which the engines of ingenuity are kept running as a consequence of the attempt to resolve existing problems and (as an inadvertent by-product) fabricate new ones.

No problem is more pressing than an existential threat. And it is arguably an existential threat that now confronts the Open Internet. What, then, is the likely response of the Open Internet? Given the disdain for regulatory countermeasures, a technological response appears to be the order of the day. But what is the nature of this response?

One possibility comes in the form of so-called Internet satellite systems. Two prominent examples of such systems are Starlink, led by Elon Musk's SpaceX company, headquartered in California, and the OneWeb Satellite Constellation, led by OneWeb, a company founded in Arlington, Virginia, and currently headquartered in London. Both initiatives aim to deploy large numbers of satellites (Starlink: 12,000 satellites; OneWeb: 650 satellites) to provide the inhabitants of planet Earth with global broadband coverage. At the time of writing, both Starlink and OneWeb have begun their launch schedules, and Starlink aims to be fully operational by 2025.

It is, as yet, unclear how these innovations—technological chimeras of both the Information Age and the Space Age—will affect the shape of the current Internet landscape. They may mark the beginning of a new kind of Internet: a Space Internet. On the other hand, they may mark a return to the global dominance of the Open Internet. Much still needs to be determined; however, inasmuch as systems like Starlink embody the values of the Open Internet and they also provide an alternative for the global community to bypass the constraints imposed by national and regional governments, then it would seem that the Open Internet is still very much a contender for global cyber dominance. Perhaps the best way to transcend an all-too-familiar set of earthly geopolitical concerns is to direct one's attention to the properties of a celestial realm.

A.2 Data Governance and Value

A.2.1 Intelligent Data Valorization

Section 5.1 discussed the use of AI techniques to support assessments of data value. This was glossed as *intelligent valorization*.

There are a number of opportunities for research in this area. One opportunity focuses on the nature of the assessment process itself. It seeks to understand the process by which data assets are evaluated for machine learning (and other) purposes. In particular, the goal is to identify the features that inform judgements of data value. Given the knowledge-intensive nature of the assessment process, the use of knowledge engineering techniques is likely to be useful for research in this area (e.g., Schreiber et al., 2000). Having said this, it is not clear that the knowledge required to complete the assessment process is available. Inasmuch as this is true, then the proposed research effort (let's call it the *epistemology of data valorization*) will be as much about the creation of new knowledge as it is about the modelling of existing knowledge.

A second area of research (not altogether unrelated to the first) focuses on issues of *volumetric optimization*. The goal here is to calculate the minimum amount of data required for specific purposes. All machine learning systems require some amount of data but understanding just how much data is required remains something of a 'black art'. In some cases, it may not be possible to determine just how much data is required in advance of the machine learning effort. Instead, one needs to monitor the progress of machine learning and judge whether an acceptable level of performance has been achieved or whether additional training is required. This feedback process is complicated by the nature of specific performance deficits. A machine learning system, for example, may prove good in some areas but not in others. In this case, the data acquisition effort needs to be aligned with the performance profile of the system at particular points in time. The question that needs to be asked in this case is not simply "How much data do I need?" but rather "What *kind* of data do I need (and what is the minimum amount I can get away with)?"

Volumetric optimization is of particular importance when it comes to personal data. This is because of the various costs and constraints associated with the collection and use of such data. In some countries, data protection legislation may impose constraints on the amount of data that can be used to train machine learning systems. And even in countries with relatively lax regulatory regimes, there may still be a requirement to minimize the amount of personal data that is collected and stored.[17]

The challenge of volumetric optimization is to determine the minimum amount of data that is required to meet a specific objective. As noted in Section 6, there is a potential parallel here with the statistical notion of power. In assessing the value of a given data asset, we need to know how likely the asset is to meet our requirements for a given objective. In the case of machine learning, the objective relates to the performance profile of an AI system. That is to say, we need to determine how much data (and, obviously, what kind of data) is required to deliver a performance profile that is 'good enough' relative to the requirements of specific applications. In addressing this issue, we clearly need a better understanding of the relationship between data and intelligence—the way in which certain kinds of data affect the performance profile of AI systems. Additional complexity comes from the diversity of machine learning algorithms currently in use. What counts as a minimally sufficient amount of data relative to, say, a deep autoencoder network, may not be the same as that required to support the same level of performance using a different algorithmic approach.

In addition to issues of knowledge modelling and volumetric optimization, the notion of intelligent valorization requires a consideration of the information used to describe data assets. In contrast to the first two research efforts—which direct their attention to the means by which data assets are evaluated—this effort focuses on the informational properties of the data assets themselves. This is important since the intelligent valorization task is likely to be of sufficient complexity as to thwart attempts at automation, at least given the current state-of-the-art. Where brains fall short, the environment must extend its reach. That is to say, if intelligent valorization exceeds the computational ambit of extant AI systems, then we will need to direct our attention to the environment in which such systems operate. In particular, we will need to consider what additions can be made to the data ecosystem in order to make the intelligent valorization task computationally tractable. The basic idea, here, is to explore ways of *enriching the data ecology* so as to transform a seemingly impossible problem into one that is much more feasible. This idea was briefly mentioned in Section 5.1, where we appealed to the notion of *data manifests*. A data manifest contains metadata about a data asset. In particular, it contains information that is pertinent to the intelligent valorization task (and thus judgements about data value). Work in this area is likely to benefit from previous research efforts, most notably efforts to develop provenance-related data models (see Moreau, 2010).

A.2.2 Comparative Analysis of Global Data Governance Frameworks

The present report provides insight into at least some aspects of China's data governance framework (see Section 5). Future work could extend the present analysis by subjecting China's data protection policies to greater scrutiny. In addition, it would be useful to have a better understanding of how China's approach to data governance compares with that adopted by other national and regional governments. Previous work has already begun to explore this issue by

[17] This is exemplified by China's *Personal Information Security Specification*, which describes a data minimization principle: "Unless otherwise agreed by the PI [personal information] subject, only process the minimum types and quantity of PI necessary for the purposes for which the authorized consent is obtained from the PI subject" (Shi et al., 2019).

comparing China's *Personal Information Security Specification* with the EU's GDPR (Gentle, 2018; Sacks, 2018a). Such studies provide an excellent starting point for more detailed analyses. Aside from the EU and China, consideration should also be given to the policy frameworks being developed by other countries.

As a means of supporting comparative analyses between global data governance frameworks, we need a means by which similarities and differences can be assessed. Although there are, no doubt, many ways of approaching this problem, one option is to rely on techniques that are used to measure the similarity of domain ontologies in computer science (see Maedche & Staab, 2002, for an example of this approach). To support such an approach, data governance policies need to be represented in the form of an ontology, using the tools, techniques, and representational formalisms developed as part of the Semantic Web initiative.[18] Such ontologies could then be compared using ontology alignment, matching, and similarity assessment techniques. The value of this approach is twofold: Firstly, it provides the basis for an ontology for the data governance domain, which is apt to be useful for governmental, commercial, and research purposes. Secondly, by using ontologies to compare data governance frameworks, we are able to compare features that are difficult to compare using other methods. These include comparisons involving semantic features. Such features include (but are not necessarily limited to) the following:

- semantic equivalence (e.g., two syntactically different policy elements turn out to mean the same thing and are thus semantically equivalent),
- semantic scope (e.g., one framework covers a broader array of regulatory domains compared to another), and
- semantic density (e.g., two frameworks differ with respect to the emphasis assigned to particular policy areas).

Comparative analyses of government policy documents come with an important caveat: while an official document records the stated policy position of a governing authority, it is not always clear that such positions reflect the reality of regulatory actions within a given jurisdiction. Consider, for example, the suggestion that China's *Personal Information Security Specification* resembles the EU GDPR (see Section 5.6). The question is whether these (on paper) similarities reflect the actual (real-world) policies adopted by the respective governments. If this is not the case, then the comparison of official government documents will tell us very little about the relationship between geopolitically distinct data governance practices. In this case, a similarity metric is probably worthless; at best it records nothing more than a tendency for regulatory echolalia.

A.2.3 (Dis)Honest Data

As noted in Section 4.4, China's vision for a SCS is predicated on the assumption that reliable records about citizen conduct can be acquired. If data collection proves to be unreliable (thereby yielding 'dishonest' data) then the system is unlikely to serve its intended purpose: If immoral behaviour is falsely reported, then deceit and dishonesty will be allowed to flourish. Conversely, if virtue goes unrewarded, or is even punished as a result of inaccurate reporting, then public support for the SCS is likely to falter.

Note that the goal of the SCS is to promote trustworthiness, honesty, sincerity, and so on in wider society. That, at least, is the stated objective of the SCS (see Section 4.1). In order for the system to work, however, it needs to track the social world in a reliable manner. It is not enough for the SCS to track behaviour, record events, compute social credit scores, mete out rewards/punishments, and so on; it must also function in a manner that secures the public's

[18] see http://www.w3.org/standards/semanticweb/.

trust. If the system itself cannot be trusted, then it is unlikely to succeed in its goal of promoting trustworthy behaviour.

This emphasis on system reliability and data credibility dovetails with recent research into what are called *knowledge machines* (Meyer & Schroeder, 2015; Smart, 2018a). These are sociotechnical systems that participate in the realization of knowledge-related processes, such as those associated with the elicitation, acquisition, and representation of knowledge. The ostensible parallels with the SCS should be clear. In short, the SCS is conceptualised as a particular kind of knowledge machine, courtesy of the fact that it trades in epistemic representations, i.e., accurate descriptions of human behaviour. To be successful, the system must possess knowledge about how a given data subject has behaved: it is not enough for the system to merely 'believe' that such and such an event has occurred; the system must instead 'know' that such an event has actually taken place.

At this point, it should be clear that the further study of SCSs provides an excellent opportunity for interdisciplinary research. Our understanding of the SCS—and its suitability as an instrument of social governance, both inside and outside China—is one that is likely to require a coordinated research effort, bringing together researchers with expertise in areas as diverse as epistemology, data science, sociotechnical systems research, artificial intelligence, political science, and trust.

A.3 Cognitive Modelling and Cyber Epistemology

A.3.1 Epistemic Safety, Algorithmic Tolerance, and Cyber Governance

Recent amendments to China's *Personal Information Security Specification* target the notion of personalized display, which refers to situations in which online information has been personalized or individualized (e.g., targeted advertising, personalized search results, and so on) (Luo, 2019). The amendments might be seen as part of an effort to minimize the negative epistemic and cognitive consequences of exposure to the Internet. This dovetails with worries about the epistemic sequelae of exposure to personalized information in the West, especially when it comes to the epistemic harms engendered by so-called filter bubbles (e.g., Government of the United Kingdom, 2019).

The notion of mandevillian intelligence provides us with a contrasting vision of the epistemic consequences of access to personalized information (see Smart, 2018b). According to the notion of mandevillian intelligence, personalized display may play a productive role in yielding collective intelligence, even though it harms the epistemic standing of individual agents. Mandevillian intelligence thus recognizes (and accepts) that certain features of the online environment may undermine the epistemic and cognitive wherewithal of individual agents, but it rejects the idea that such features are bereft of any sort of epistemic or cognitive benefit. Filter bubbles may thus pose a genuine epistemic threat to the individual agent, but this does not mean that such threats extend to communities of such agents. Indeed, in some cases, it may be that collective forms of cognitive success are predicated on individual cognitive shortcomings—that the (cognitive) virtues of the many are predicated on the (cognitive) vices of the one.

Mandevillian intelligence presents a significant challenge to prevailing views about the epistemic and cognitive value of certain kinds of online phenomena. It also comes with important implications for cyber governance. If our aim is to protect the epistemic integrity of individual citizens, then filter bubbles are apt to be perceived as a threat and in need of some sort of technological or regulatory remediation. Suppose, however, that in addition to being concerned about the rights, entitlements, and privileges of the individual citizen we are also concerned about the cognitive well-being and epistemic standing of agent communities. According to the notion of mandevillian intelligence we now confront a dilemma: What is more important (or valuable) to us: the rights of the one or the rights of the many? If it is the former, then a regulatory response is required; if it is the latter, then we might be inclined to take a 'hands-off' approach and leave things as they are.

How does this play out in a geopolitical context? Given China's socialist leanings, we might expect it to focus on the collective good. If so, then perhaps it ought to take a hands-off approach when it comes to personalized display. As evidenced by recent government publications, this does not appear to be the case.

Inasmuch as neo-liberalism is concerned with the rights and responsibilities of individuals, then we might expect Western democracies to be less tolerant of algorithms that pose a threat to individual epistemic standing. (This certainly appears to be the case in Europe.) On the other hand, liberal democracies typically express support for laissez-faire, free-market approaches. Inasmuch as we accept the notion of mandevillian intelligence, then there might be some merit to such approaches: From a mandevillian perspective, a preoccupation with individual freedom and market autonomy might be the best way of securing the collective cognitive good.

A.3.2 Vice, Virtue, and the Shortcomings of Social Credit

The notion of mandevillian intelligence is inspired by the work of the Anglo-Dutch philosopher and political economist, Bernard Mandeville. It calls for a relativistic conception of cognitive/epistemic virtue that distinguishes between individual and collective levels of analysis. This dovetails with Mandeville's views about the relation between individual conduct and the socio-economic good. In *The Grumbling Hive*, published in 1705, Mandeville describes a bee community thriving until the bees are suddenly made honest and virtuous. Without their desire for personal gain, their economy collapses, and the remaining bees go to live simple lives in a hollow tree, thus implying that without private vices there exists no public benefit.

The moral of the story, if true, presents a potential problem for China's SCS. The SCS, recall, is motivated by the attempt to instil virtue at the level of individual citizens (see Section 4). If Mandeville's bees are anything to go by, however, this is a mistake. A society of saints no doubt comes with certain benefits, but whether those benefits exceed the sort of benefits that might be obtained in a less virtuous society is unclear. In any case, if private vice plays a positive functional role in securing public benefits, and what we seek to maximize are public benefits, then what is the basis for punishing those who have helped us to achieve our socio-political objectives? If vice and virtue are (in effect) the yin and yang of the socio-economic good, then we cannot have one without the other. In this sense, the SCS poses as much a threat to socio-economic success as it does a route to political security. As an example, China's approach to illegal downloading and pirated content may have been instrumental in supporting the growth of Internet penetration. As noted by Fabre (2018):

> . . . the lax regulation of digital content in terms of intellectual property rights strongly enhanced the penetration and the use of [Internet] devices, with free and easy access to music, books and movies.

This points to one of the potential virtues of what might otherwise be regarded as an individual vice, with Internet penetration yielding a mix of both economic benefits and opportunities for technological innovation (see Section 3.4.2).

Here we see how the notion of mandevillian intelligence might be used to challenge Chinese government policy by confronting it on its own philosophical turf. In other words, we accept the Chinese leaderships' basic commitment to systems-level thinking and a scientific approach to social engineering, but we then proceed to show how this approach mandates a consideration of forces and factors that undermine the logic of social credit policies. This is, perhaps, a much better way of critiquing the SCS, as opposed to lambasting the Chinese leadership from a politically and ideologically remote soapbox.

A.3.3 Cultural Network Analysis

Cultural network analysis (see Sieck, Rasmussen, & Smart, 2010) could be used to model the collective cognitive bases of Chinese policy. This promises to deliver a better understanding of China's policy agenda, help us identify points of influence, and enable us to predict China's response to international action. This could then be used to inform responses to China's domestic and international policies.

This approach assumes that nation states can be modelled as (collective) intentional systems that can be understood in the same way that we understand the behaviour of rational agents. By adopting an intentional stance (see Dennett, 1987) towards nation states, we can ask questions such as: What are China's beliefs and desires? What is the doxastic justification for specific actions? What rationalizations does China rely on? How does China perceive the world?

A virtue of this approach is the way it helps us understand the (cognitive) impact of global events. Consider, for example, the claim that China 'perceives' an open and ungoverned Internet as a potential threat to political and social stability. Such concerns are undoubtedly reinforced by various revelations and scandals, such as the Snowden revelations, the Cambridge Analytica scandal, and cyberattacks by foreign actors. From China's perspective, these scandals (and the ensuing critique of large technology companies, such as Facebook) are apt to legitimate the need for an authoritarian approach to cyber governance. In a sense, then, Chinese policy is, at least in part, forged in a Western furnace. It is the apparent shortcomings of the Open Internet, and the Western response to such shortcomings, that rationalizes the need for interventions that seek to safeguard the interests of individual Internet users and protect national security.

A.3.4 (Collective) Cognitive Dissonance

Is China's regulatory approach internally consistent? Is it compatible with the CCP's strategic objectives? The following are some areas where doubts might be raised.

Inclusive Globalization: The term "inclusive globalization" arises in respect of China's Belt and Road Initiative (see H. Shen, 2018). It refers to China's attempt to preserve economic globalization, possibly in response to Donald Trump's "America first" mantra.[19] This includes support for Alibaba's efforts to construct an Electronic World Trade Platform (eWTP), which is intended for small and medium-sized enterprises in developing countries.[20] It is unclear how China's apparent support for economic globalization and free trade under the banner of "inclusive globalization" tallies with China's protectionist approach to data governance.

Data Rights: In regard to data protection, the Chinese leadership appears keen to uphold an individual's "Right to Erasure" (or right to be forgotten). It is, however, difficult to see how this 'right' can be preserved if one is simultaneously trying to use online behaviour for the purpose of social credit calculations.

User Consent and Extended Use: China is keen to restrict the sharing and use of data in ways that go beyond the bounds of the explicit consent mechanisms used

[19] It should also be remembered, of course, that the America first policy may reflect a policy response to China's support for indigenous innovation and domestic technology companies. In essence, both China and the US may be moving towards a protectionist policy, while simultaneously espousing the virtues of globalization.

[20] See https://www.ewtp.org/.

by Internet services. It is unclear how Chinese companies can comply with this legislative constraint if they are also legally bound to share data with government authorities (and other agencies) for the purpose of building a SCS. In short, the data requirements of the SCS appear inconsistent with China's approach to data protection. The *Personal Information Security Specification*, for example, stipulates that personal data cannot be used for "extended functions" or "secondary uses" (i.e., data processing activities that fall beyond the bounds of a consent agreement). Arguably, however, the use of personal data for the purpose of computing social credit scores counts as a form of extended function or secondary use.

Personal Display: China's data protection policy attempts to safeguard the interests of its citizens when it comes to personalized display and personalized search. But this policy does not seem to extend to the Internet as a whole. Inasmuch as all search results in China are 'filtered', shouldn't citizens be made aware that they do not have access to the 'bigger picture'? Courtesy of censorship, Chinese citizens are still presented with a filter bubble—or 'propaganda bubble'—of sorts. This is one example where China's cyber/data governance policies could be seen to lack internal coherence and consistency.

From a collective cognitive standpoint, these examples point to a degree of cognitive inconsistency or cognitive dissonance (Cooper, 2007) at the heart of China's cyber and data governance policies. The further analysis of these 'dissonant states' may shed light on the direction of future policy-making efforts.[21]

A.4 Understanding China

A.4.1 Sino Situation Awareness

How advanced are China's plans to implement the SCS? Has there been any attempt to empirically evaluate the effectiveness of specific SCS systems, such as the local pilot projects discussed in Section 4.2.2?

Progress on these issues is hampered by the difficulty in accessing studies by the Chinese scientific community. The absence of empirical evaluations of Chinese SCSs in Western academic journals may mean that such studies have not been undertaken. Perhaps, however, such studies *have* been taken, but the results are reported in Chinese academic journals and are thus less accessible to a Western audience.

Future work would benefit from the creation of something akin to a 'Chinese Google Scholar'—a system that enables Western researchers to search Chinese scientific repositories. In support of such a system, cloud-based machine translation services could be used to generate English language annotations of Chinese resources. Note that there is no requirement for such a system to copy and translate actual articles; it simply needs to provide some insight into what articles exist. This can be done by working with the metadata associated with academic publications.

In addition to Chinese scientific journals, it would be useful to extend the reach of such a system to other types of resources, such as media reports, official government plans, articles by civilian and military scholars, technical publications, and other trustworthy online resources. Access to such resources would be invaluable in furthering our understanding of China's plans

[21] Also note that the complexity of China's policy framework is a potential problem (for them). It risks introducing inconsistencies, but it also destabilizes the system of beliefs that are used to sanction particular response outcomes. The more beliefs one has about an issue, the more opportunities one has to switch to an alternative policy position. See Smart et al. (2010) for more on this.

for the SCS, as well as other areas of strategic importance, such as military applications of AI technology. As a point of departure for such efforts, it is worth considering the China National Knowledge Infrastructure (CNKI), a Chinese-language database comprising academic journal publications, newspapers, dissertations and technical papers.[22] The CNKI is a national information construction project, which was launched in 1996. It is currently led by Tsinghua University with support from the Ministry of Education, the Ministry of Science, the Propaganda Department of the Communist Party of China, and the General Administration of Press and Publication.

A.5 Artificial Intelligence

A.5.1 Autonomic and Artificial Governance

Throughout the present report, we have seen how China seeks to use technology as an instrument of social governance and political control (see, for example, Section 4.1). This raises issues about the extent to which socio-political processes could be realized via mechanisms that are created, managed, and maintained by AI systems. Mechanistic approaches to understanding social phenomena are well-documented in academic literature (Hedström & Bearman, 2009; Hedström & Ylikoski, 2010; Ylikoski, 2018), but such approaches have seldom been applied to the political domain. Future work in this area could explore the way in which the Chinese leadership seeks to use technology for political purposes, thereby establishing sociotechnical mechanisms that enforce the values, beliefs, and precepts of a ruling elite. This would serve as an interesting focus area for the proponents of mechanistic approaches in the social sciences. In particular, it provides a nice opportunity to extend the reach of mechanistic analysis to the realms of political science and the study of political phenomena.

A.5.2 Artificial Economics

China sees AI as a route to economic growth. In particular, policy documents speak of the "deep integration" of AI into the economy (State Council, 2017). China's vision is essentially one of using AI to improve certain kinds of business processes, such as the use of automation techniques to improve manufacturing processes.

Suppose, however, that we take a step back and look at economic systems from the standpoint of a mechanistic approach. In other words, let us think of economic systems as something akin to a machine whose various routines (i.e., economic processes) are realized by the operation of (in this case) economic mechanisms (see Marchionni, 2018). A major constituent of such mechanisms are entities that function as economic actors. Such entities may be individual humans or corporations comprising individual humans, but this detail need not concern us here. What is important, for present purposes, is the idea of economic phenomena (e.g., economic events and processes) being realized by a material fabric that includes human agents as constituent elements. In other words, human agents are the material components of economic mechanisms.

Now let us consider the status of AI systems relative to such a vision. China's vision of deep integration sees AI systems as altering the functional properties of processes that exert an impact on economic variables. By improving the efficiency of manufacturing techniques, for example, productivity is improved, along with the prospects for economic growth. As it stands, however, AI systems are, at best, weakly integrated into economic mechanisms. For the most part, AI systems are not constitutively relevant to the realization of economic phenomena; rather, their role is one of mere causal relevance (see Craver, 2007, for a discussion of the difference between causal and constitutive relevance). That is to say, AI systems exert a causal influence on

[22] See https://www.cnki.net/.

economic phenomena, but they do not constitute those phenomena, in the sense of functioning as the components of economic mechanisms.

The shift from causal to constitutive relevance is accompanied by a corresponding shift in the way we think about AI systems. We need to ask ourselves what would be required in order for AI systems to function as *bona fide* economic agents. One answer to that question comes in the form of what is called *artificial economics* (Smart, 2020). According to this proposal, the status of AI systems as economic agents rests on a dual capacity for both production and consumption. In other words, AI systems (*qua* economic agents) function as *economic prosumers*—they are not just the producers of goods and services, they are also the consumers of goods and services. It is this combined producer/consumer (or prosumer) capacity that then enables AI systems to function in more or less the same manner as their human counterparts, at least when it comes to the realization of economic phenomena. This is important, for such capacities may serve as something of a technologically-oriented fix to the limits confronting capitalist economies, especially when it comes to issues of economic growth and capital accumulation. Ultimately, there are only so many things that humans can consume, and there are only so many (human) consumers that can be accommodated on planet Earth before the whole socio-economic shebang topples over the edge of an ecological precipice. Some commentators have suggested that such constraints portend the demise of capitalism, or at least the end of its expansionist agenda (Kovel, 2007). From the standpoint of artificial economics, however, there is still all to play for. In the short term, capitalism stands to benefit from the forms of ‘deep integration’ alluded to by the Chinese leadership. In the longer term, a much deeper form of integration—embedding AI systems in economic mechanisms—may help to free capitalism from the constraints imposed by its human creators.

A.5.3 Cyber-Leninism, Free Energy, and the Mind of Society

China’s leadership embraces a positivist view of social reality based on systems theory, which holds that society can be understood and engineered through a holistic, scientific approach. This dovetails with mechanistic approaches in the social sciences, such as those favoured by the proponents of analytical sociology (see Hedström & Bearman, 2009). The Chinese leadership also seeks to use technology to improve their ability to understand and predict society. Creemers (2017), for example, suggests that technology plays an important role in improving the “legibility” of Chinese society. In particular, he suggests that “the Party-state seeks to deploy technology in a manner that renders society legible and predictable” (p. 88).

From an AI perspective, this emphasis on understanding, prediction, and control strikes a chord with recent work in theoretical neuroscience and deep machine learning. Contemporary models of brain function appeal to the idea that the biological brain is a hierarchically-organized prediction machine that acquires a so-called generative model of the sensorium as a means of reducing prediction error, which is an information-theoretic isomorph of statistical free energy (see Clark, 2016). Generative models are also a feature of recent work in deep machine learning. In particular, some strands of deep learning research emphasize the importance of generative capacities, as opposed to the more traditional emphasis on discriminative capacities. This shift in emphasis is reflected in the goal of machine learning. From a generative perspective, the aim of a machine learning system is to generate (or recreate) the bodies of training data that it is exposed to. This leads to the acquisition of a generative model that embodies the (hidden) causal structure of whatever processes are responsible for patterns in the training data.

From a biological standpoint, the success (or fitness) of a generative model inheres in its capacity to predict sensory inputs. A good generative model is thus able to anticipate the flow of sensory information, courtesy of the fact that its learning regime is driven by a computational imperative to reduce prediction error (or free energy). To minimize prediction error, a generative model must be able to generate (from the top-down) the pattern of sensory inputs that will ensue at future points in time. This requires a knowledge of how the future will unfold, which is itself rooted in an understanding of how causal forces in the world interact to yield moment by moment changes in the sensory flux. This is, in fact, the essence of understanding: If a generative model succeeds in predicting the world, then it understands the world; if it fails to predict the world, then prediction error will ensue and there is more learning to be done (see Clark, 2013).

Unless that is, the generative model exerts some *control* over the sensory environment. If the generative model is able to effect changes in the world (as is the case with mobile, embodied agents, like ourselves), then it can alter the structure of the world so as to simplify the learning task, e.g., by making the world more predictable. In fact, if the control is absolute, in the sense that every change to the world yields a predictable response, then the generative model does not need to learn about the causal structure of the external world at all. For the generative model is, in this case, the cause of its own inputs. All the relevant causal structure is thus *internal* to the generative model itself, and all the model needs to do is learn how to predict its *own* actions. . . that is, it needs to learn how to predict, and thereby expect, *itself*.

We thus arrive at a compelling vision of the mechanistic underpinnings of both natural and artificial intelligence. Such intelligence is born of the need to predict the future and thereby minimize free energy or prediction error. Residual free energy drives learning, which, in the longer-term, yields generative models that embody the causal structure of the sensory environment. With predictive success comes understanding—understanding is, in short, the capacity to confront the world while keeping free energy at a minimum. For active systems, there is an additional wrinkle. Active systems can exert control over the nature of their sensory inputs and thus maintain free energy at an acceptable level. The result is that understanding the world is, in part, a function of one's capacity to control the world.

What, then, of the Chinese leadership's attempt to deploy technology in a manner that renders society legible and predictable? The goal here is to understand a particular environment, namely, a human social environment, or Chinese society. The advent of the Internet provides an opportunity to monitor the 'sensory' data that emanates from that environment; in essence, the Internet provides opens a 'window' onto the human social world. For the sake of illustration, let us assume that the Internet forms part of a sophisticated surveillance infrastructure that enables government bodies to observe the ebb and flow of social events at a variety of spatial and temporal (and, perhaps, social) scales. Social events are thus rendered in the form of prodigious quantities of digital data. Making sense of that data requires an advanced form of intelligence. Imagine, then, that we employ the services of a deep learning system whose aim is to acquire a generative model of the relevant sensory environment. The sensory environment, in this case, is the human social world—society itself. The model to be acquired is thus a particular kind of generative model: not a model of the traditional physical world (the world that is revealed to us via conscious experience), but a model of the social world—a socially-oriented generative model, or a generative model of social reality.

The goal of our system, recall, is to understand society. It does so by changing its internal architecture to minimize free energy. Free energy, in this case, reflects the mismatch between the predictions of a multi-level, socially-oriented generative model, and the patterns revealed by digital renditions of the social domain. If learning succeeds, then the generative model will embody the causal structure of the social environment: it will understand society, courtesy of its capacity to predict society.

Societies, however, are unruly beasts, and it is doubtful whether any AI system, regardless of its sophistication, could predict the course of social events. This is where issues of control and governance come into play. In particular, authoritarian modes of social governance seek to exert control over the social environment, and, in doing so, they make the social environment easier to predict and thus easier to understand. The best way to predict the future is to create it, and this is arguably what authoritarian forms of social governance strive to achieve.

The upshot of all this is an area of research that seeks to explore the application of deep machine learning to the human social environment. It is an area of research that speaks to China's interest in using AI to understand and predict society. It is, moreover, an area of research that brings together many of the issues encountered throughout the present report. By seeking to create generative models of the social world, we are forced to confront issues that straddle the domains of social governance, the Internet, AI, and personal data.

Research in this area is timely, for it is only with the advent of deep machine learning and access to voluminous quantities of social data that the notion of socially-oriented generative models becomes feasible. It is also an area of research that dovetails with China's political and technological ambitions, especially when it comes to the use of AI to support the intelligentization of social governance. Finally, it is worth noting that the sort of AI system envisioned here will only work in a particular socio-political context. It requires a society with an expansive surveillance infrastructure and a government-mandated imperative to exploit personal data for political purposes. It also requires a robust commitment on the part of a national government to support the learning objectives of the AI system—to make society predictable (and thus learnable) by imposing constraints that limit the degrees of freedom for social action. China is arguably the perfect setting for such a system; it is perhaps the only country in the world where such a thing could be achieved.