LIBERAL ARTS AND HUMAN SCIENCES
CENTER FOR HUMANITIES
VIRGINIA TECH.

Mellon
Foundation

Tech for Humanity Case Studies

The Smart Big Brother

George Orwell's novel *1984,* published in 1949 following World War II and rise of the Iron Curtain gave a fictional account of a society completely succumbing to the control of the state. While Orwell's work was fiction, the methods examined in the novel were quite close to the reality experienced by many Soviet citizens from the 1930s until the end of the Cold War in 1991. The mechanisms of control in *1984* illustrate a socio-technical system whereby the state gains total physical and psychological control over its citizenry. The book has become standard required reading in literature classes and has become an often-used point of discussion in sociology, political science, and even science and technology studies. The powerful plot was even used as a metaphor to represent control in other sectors, most notably Apple Computer Inc released an advertisement in the year 1984 portraying software and hardware giant IBM as the monolithic big brother from the novel being overthrown by sledgehammer wielding woman. As big brother on the screen states:

> *"Today, we celebrate the first glorious anniversary of the Information Purification Directives. We have created, for the first time in all history, a garden of pure ideology—where each worker may bloom, secure from the pests purveying contradictory thoughts. Our Unification of Thoughts is more powerful a weapon than any fleet or army on earth. We are one people, with one will, one resolve, one cause. Our enemies shall talk themselves to death, and we will bury them with their own confusion. We shall prevail!"*

The sledgehammer wielding woman tosses the tool towards the screen broadcasting big brother destroying it in a vibrant display of destruction and shocking the audience out of their induced stupor. As the commercial comes to an end the screen reads "On January 24th, Apple Computer will introduce Macintosh. And you'll see why 1984 won't be like *1984*."

Orwell's novel, and Apple's advertisement serve as both the intellectual and technical origins of the limitless possibilities that lay ahead for the future of surveillance. There is no doubt that Apple wished to foster a nascent personal computer enterprise that extended computational resources beyond mainframes to the general-purpose, publicly accessible computing, but touching on the 1984 allegory would come back to haunt it decades later as its and other mobile devices have become the technical tools in service of Orwell's dystopian vision.

Mobile phones became popular in the 1990s, and mobile computing devices rose to prominence with the BlackBerry 5810 in 2002. Yet, it was the advent and release of Apple's iPhone in June 2007 that would change how we interact with the Internet and usher in a rush to digitize our lives in ways previously unimagined. Soon users all over the world carried mobile devices with enormous volumes personally identifiable data everywhere they went. These devices combined with advances in algorithms for search, tracking user behavior, and hardware advances that added functionality through suites of sensors including accelerometers, cameras, microphones and more have generated enormous markets, fostered amazing efficiencies, and enabled the creation of a dystopia beyond even Orwell's wildest dreams.

Although users and liberal democracies saw these devices as tools of empowerment, authoritarian regimes saw threats to the status quo. The perception of technology as a threat among authoritarian regimes began to gradually shift towards opportunity in the mid-2000s. While many around the world were shocked by the 2013 revelations about American surveillance activities revealed by Edward Snowden, many states also saw new potential domestic applications that could be implemented in service of the state.

No state has spent more time and effort leveraging advances in technology to facilitate domestic control then China. Beginning in the late 1990s China began to identify digitally connected technologies as a threat to regime and societal stability. By the early 2000s China had begun enacting regulations to force technology firms to adhere to Chinese state information and security policies. By 2002 China had begun blocking western technology firms such as Google and began to "force" companies that wished to remain in their enormous market to comply with Chinese information and technology transfer requests. By 2006 China had implemented its first version of what has become known as the "Great Firewall of China (GFC)." The GFC is a mix of laws, policies, and technologies that regulate/control China's domestic Internet. Much of this regulation censors around censorship of external content, or content that does not conform to Chinese Communist Party information rules. Concurrent to the development of the GFC China also developed what it coined as the Golden Shield Project (GSP). The GSP is a project of the Ministry of Public Security. In establishing the GSP the Ministry set out the following articles stipulating its mission:

> *Individuals are prohibited from using the Internet to: harm national security; disclose state secrets; or injure the interests of the state or society. Users are prohibited from using the Internet to create, replicate, retrieve, or transmit information that incites resistance to the PRC Constitution, laws, or administrative regulations; promotes the overthrow of the government or socialist system; undermines national unification; distorts the truth, spreads rumors, or destroys social order; or provides sexually suggestive material or encourages gambling, violence, or murder. Users are prohibited from engaging in activities that harm the security of computer information*

> *networks and from using networks or changing network resources without prior approval.*

In essence, the GSP picked up domestically at the edge of where the GFC ended in blocking and censoring foreign content. The GSP sought to surveil and police nearly all interactions on Chinese domestic networks. Since 2006 the technologies of mass censorship and surveillance in China have only expanded. As China's aggressive position towards western firms intensified, many of these firms began to leave the Chinese market. Yet in their wake they left behind large numbers of trained software and hardware engineers who used information gleaned from technology transfer programs and outright theft of trade secrets to foster an emergent domestic technology sector. By building on the backs of billions of dollars' worth of Western Trade secrets and substantial state investment Chinese firms were able to rapidly advance in the fields of artificial intelligence and networking. In the span of 20 years China developed itself into one of the leading technological competitors of the United States. Technological development coupled with large scale state investment has enabled China to become a dominant global player in networking and software distribution. Distribution networks are further solidified and locked-in through Chinese financing for third party nations.

Chinese networked infrastructures have been putting all the pieces of the technological puzzle together since 2014 to create the ultimate surveillance state. Building on a concept first developed by Jeremy Bentham of a perfect prison in which a guard can see all the prisoners, but the prisoners cannot see the guard, the Chinese GSP and GFP were the initial phases in the development of a digital panopticon. Maximizing societal control prior to the digital/networked era was expensive and required large scale human infrastructures. For instance, at the height of the East German secret police the Stasi 91,000 regular employees and 174,000 informal employees constituting approximately 1 state security officer for every 54 citizens were employed for state internal surveillance purposes. Human surveillance and security networks are expensive and difficult to manage. Moreover, such networks while attempting omnipresence have numerous gaps, notably within the privacy of an individual's home or while traveling to third party states.

To address these shortcomings the Chinese state is leveraging omnipresent mobile devices and applications pioneered by the likes of Apple, Samsung, Huawei, Google, and others in combination with AI technologies ranging from facial recognition to anomaly detection. The result has been the creation of an increasingly comprehensive social credit system (社会信用体系). The Chinese social credit system combines the surveillance capabilities of dozens of Chinese technology firms ranging social networking, network infrastructure, and hardware firms into unified and integrated state managed databases with AI models assessing behavior and making predictions. Each of these companies provides products and services that are used by Chinese citizens daily on a plethora of digital devices. Each of these devices transmits and stores personally identifiable information including user behavior. This amalgamation of devices and information extends well beyond public spaces into the intimate private spaces of users is augmented by an immense network of additional surveillance technologies increasingly empowered by AI technologies. Among the most prominent of these is the national system of networked surveillance cameras, SkyNet. The system, sharing the same name as the AI from the Terminator movies, consists of more than 200 million CCTV cameras in tandem with advanced facial recognition technologies to provide near total national coverage within urban areas.

Once all these systems were combined and the databases synchronized the result was a societal level panopticon capable to reaching quite literally into the minds of Chinese citizens. China began to emphasize the use of technology in societal surveillance and has leveraged the global rhetoric surrounding terrorism to tailor its approach towards targeted sub-populations. Uyghur populations in XinJiang province have been subjected to extreme levels of surveillance. Targeting of Uyghur populations is ostensibly in response to a series of terrorist attacks from 2010 to 2017 that sought regional autonomy among Uyghur populations. Several of these attacks, including a 2013 use of a vehicle as an incendiary device in the Famous Tiananmen Square in Beijing, brought the full force of the Chinese surveillance state down on Uyghurs both in XinJiang and globally.

Many initial attempts to utilize AI to target Uyghur populations were weak and ineffective. Chinese software firms simply did not have enough data to develop and train effective models to target and identify groups within the population. Firms, often owned and staffed by former employees of Western software firms began developing or co-opting AI models for use in China. As Western tech firms were slowly eliminated from domestic markets these new start-ups rose to take their place. By building close and ongoing relationships with the Chinese state these firms were able to exploit the available market and grow rapidly. As they grew, they incorporated an increasing array of capabilities enabling the collection of new data types. These data types included geolocation data, health data (including heart rate, steps, exercise time), purchasing data (including what, where, when), bill payment data, social media data (social networks, political, entertainment, interests, etc.) and multiple other data types. Often datatypes were segmented by device or application on the device. In the early stages of state surveillance individuals were required to put certain applications on their devices. These applications shared various data types with authorities and helped build a pattern of life.

Early attempts to synthesize this data were deemed insufficient to counter the challenges associated with managing an entire sub-population of millions of individuals. As a result, the state began requiring entire families to receive state mandated physicals which began the process of deepening the data repositories on individuals. In such physicals individuals were measured, had multiple photographs taken of them from different angles, gave blood samples for genetic testing and more. With this new repository of data linked to the social, financial, and other behavioral information, the state was able to build robust digital profiles for individuals within the sub population. As a result, they could track any perceived or AI identified anomalous behavior. An example of pattern of life tracking:

Normally take a walk in the morning? Didn't go for a walk today? Are you plotting a terrorist attack now?

The AI is unable to distinguish between normal and abnormal behavior on small sample sets. So an individual with the flu, a cold, or who just didn't feel like taking a walk in the morning because they were up late studying or watching a TV show now becomes and object of state surveillance.

As individuals progressed throughout their daily lives, they were increasingly subjected to enormous levels of surveillance ranging from the collection telemetry from digital devices to facial scans conducted at check points, or on cameras in stores or at intersections. In some extreme cases the state even mandated the installation of devices within homes to monitor behavior. Where once the state leveraged people against people, this human surveillance network is now augmented with digital technologies such that nowhere is immune to the watchful eye of the state. Since achieving near ubiquity in 2017 and only improving in its breadth and depth of reach since, there have been no further terrorist attacks conducted by Uyghur populations. Surveillance combined with various forms of internment in prisons, reeducation centers, concentration camps and other facilities, gave the Chinese state total visibility on the Uyghur sub population in Western China. Moreover, novel surveillance techniques combined with international agreements facilitated in depth access to the activities of Uyghurs in diaspora communities in countries around the world.

Adding to the surveillance networks are the AIs that operate behind the scenes. Non-auditable models "predict" who is likely to be a "terrorist" and then government officials arrest or detain persons identified by the AI. This form of detention is most comparable to pre-crime detentions for something that may never happen. Moreover, individuals detained might never fully understand what led to their detention. Time and again Uyghur populations have been subjected to intrusive surveillance techniques that offer little in the form of transparency or equality. What's more, the same technologies being employed against Uyghur populations are increasingly available for export to other regimes around the world.

What made Orwell's dystopian novel a reality? Algorithms, hardware, and data repositories that began as voluntary eventually morphed into mandatory instruments in the service of an intrusive state. The iPhone consolidated cameras, music, mobile browsing, and other behaviors into a simple handheld device that has now become ever more powerful and capable of transmitting data. It initiated a class of devices, software, and data, that has streamlined and made our lives increasingly efficient. These devices and algorithms have provided enormous utility in the form of convenience and entertainment. Technologies that we now use daily are integral to new markets with billions in revenue and have created millions of jobs. Yet they have also empowered states like China to become the smart big brother to all her citizens.

Discussion Question #1

Fiction often serves as a lens through which we can see the future. What novels or movies have you read or watched that have made you think about the potential future implications of technology? Did these works of fiction make you feel empowered or powerless?

Discussion Question #2

George Orwell used fiction a tool to engage in political commentary on the behaviors of the state in controlling society. Do you think his dystopian future has come to pass? Are we beyond the point of no return? How can you as a citizen in a liberal democracy address the use of technology by states for surveillance purposes?

Discussion Question #3

The case analysis above links the creation of new digital communications devices, specifically the advent of the iPhone to the use of technology by states such as China for surveillance. Is this linkage fair? Would we have been better off without such devices? Would it have been better to think through the potential impacts of such devices on humans prior to their release? Although Apple claimed it was fighting IBM’s big brother, Apple is now one of the world’s wealthiest companies. Did Apple inadvertently become the big brother it sought to destroy?

Discussion Question #4

The case above focuses largely on the use of technology by China for surveillance purposes. All the information presented above has been documented in numerous case reports, interviews with members of ethnic minorities and other targeted populations, and technical means of analysis, including network collection and satellite imagery. China justifies its approach to surveillance as a necessary evil to maintain a stable society. What do you think about this justification? What is the proper balance between societal stability and security and other rights and privileges? What role should technology play in such considerations? What role should the companies that develop technologies for potential use in surveillance applications play?

Discussion Question #5

Artificial intelligence algorithms are often the key ingredient in linking disparate data sources. AI as a family of algorithms provide novel insights that can pave the way for new tools and services and often provide understanding of events or individuals obscured through data glut or complexity. What do you think the role of AIs are in facilitating the development of a robust surveillance state? How should such AIs be governed? Should they be governed?

Discussion Question #6

If some states govern AI usage and others do not what are the ramifications of such an imbalance? Should all states push the boundaries of AI development, or should there be limitations? If there are limitations how should states negotiate such limitations?

Discussion Question #7

AI is a critical component in surveillance systems due to the role it plays in aggregating and analyzing large volumes of data. Most AI systems are designed and developed by humans. To what extent do humans imbue their creations with their own heuristic biases? What are the implications of these biases on the resultant algorithms, data analyses, and human security more broadly?

Discussion Question #8

The Chinese state has played an outsized role in the development of artificial intelligence for both private and public use. The influence of the Chinese state has extended from financial resources to fund research and development, purchase of new software and hardware goods and services, laws and policies that shape the market, and foreign intelligence operations that gave domestic firms an advantage they might not otherwise have had. Know that China has played such a large role in the development of surveillance AI, should the United States and other powers do the same?

Reflecting on the Smart Big Brother

There have been many neigh Sayers who have spoken against the development of artificial intelligence. In an open letter entitled: “Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter” signed by Stephen Hawking, Elon Musk and dozens of other experts on AI write:

> *The potential benefits are huge, since everything that civilization has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools AI may provide, but the eradication of disease and poverty are not unfathomable. Because of the*

> *great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls.*

The case above illustrates one avenue of the pitfalls potentially arising from AI development and use. Yet the Chinese case is not unique. Nor is the case unique to authoritarian regimes. The misuse of technology leading to potential pitfalls abounds not only in popular science fiction and in non-democratic states. Rather, the realization that technology can be miss-used is unfortunately a consistent reality repeatedly demonstrated. The interweaving of software, hardware, networking, and AI can lead directly to issues pertaining to transparency, justice, accountability, and liberty.

Transparency: AI poses a unique challenge to transparency. As AIs bring together data derived from devices and networks and processes this data using a variety of techniques its output in the form of decisions or analyses can frequently seem arbitrary and capricious. In the case above many Uyghur did not know or understand why they were being targeted by the AI. Many of the police and security forces similarly did not know why they were arresting certain individuals beyond that a system which they do not understand told them to do so. As AI is increasingly used by the state it will become ever more important to open the black box processes of decisions arrived at by algorithms. By forcing AIs to explain their actions, to demonstrate how a decision was arrived at, and making the whole process auditable transparency is enhanced and confusion is minimized. Simply putting in place transparency will not solve the miss-use of AIs by states or other actors, but it will help us to understand actions more clearly.

Justice: AIs are not independently just. They do not have moral or ethical reasoning capacities. Rather they are analytical machines that are a simulacrum of the systems in which they were developed. AIs developed in societies which pay closer attention to issues of justice are more likely to produce just solution sets. AIs developed in societies with high levels of discrimination, inequality, and bias are likewise likely to reflect those qualities. The AIs used in the Chinese surveillance system represent the beliefs and preconceived perceptions of the dominant Han population. The result is that the systems become increasingly unjust as they are applied to individuals outside of the primary or majority culture. Similar outcomes in the United States and Europe have been well documented. In Western states such as the United States predictive policing AIs have led to higher arrest rates within minority communities. Audits reveal that these AIs foster a self-reinforcing environment or a feedback loop that criminalizes certain neighborhoods and populations. The case above makes clear that in reaction to terrorist attacks in China the state has reacted harshly to curtail all potential expressions that diverge from the party line. The result is a system that was both embedded with injustice and that through continuous feedback has become less just over time. China is unlikely to alter the trajectory of its AI development, yet its use AI in ways that undermine justice can serve as a lesson for the use of

AI in policing, and intelligence in other states. Ensuring justice is pervasive is time consuming and requires often contending with uncomfortable issues such as bias, or grievances within the broader culture, political, or societal structure.

Accountability: AIs are code-based models that leverage increasingly advanced computational infrastructures. These models are nearly all human built and designed. Anthropomorphizing AIs and asking if an AI is accountable is a poor substitute for asking whether the humans who use or design the AI are accountable. It is all too easy push decisions off to an automated system. The assignment of accountability away from the human to the machine absolves the human of moral and ethical responsibility. This absolution of responsibility makes it possible for humans to engage in ethically and morally repugnant manners. Police officers can tell Uyghur detainees they are being arrested for being a terrorist threat with little regard to the reality of a given situation. They can do so because they assign responsibility to a machine. Yet, in the end when a machine is deemed the accountable party, no one is accountable. This is because there are no meaningful consequences that a machine can suffer for making the wrong choice or giving faulty analysis. Even if a machine is deemed incapable of providing accurate information and subsequently deactivated, this deactivation does not constitute accountability. The machine neither knows whether it is right or wrong. The machine will feel no pains of guilt for the children of a single mother sent to an internment and re-education facility who end up in the care of the state. AI accountability must resolve to both the programmers and the operators. Fundamental mistakes in the programming of an AI would imply developer accountability. Mistakes in the utilization of an AI would make the operator accountable. At times there will be a balance between the accountability of both parties.

Liberty: Liberty is defined by the Oxford English Dictionary as – (1) the state of being free within society from oppressive restrictions imposed by authority on one's way of life, behavior, or political views. (2) the power or scope to act as one pleases. In many ways the devices and services that increasingly define modern life enable liberty. They place a world of information at our fingertips and provide us the opportunity to participate in a global information space where we can share our ideas. Yet, these same devices of empowerment can equally be tools of repression. Devices which empower us with information and communication are increasingly being used as the wardens of digital prisons empowered by AIs and large volumes of data. The case above illustrates how taken to an extreme the devices which enable efficiency, communication, and so much more can also be turned against their users. They become the eyes of the state in our most intimate spaces. These devices serve as beacons broadcasting our innermost thoughts, fears, wants, and desires. At times, data generated through passive and active interactions in increasingly complex digital ecosystems provide insights into our lives that we, our friends, and even our family might not fully comprehend. In the hands of corporations this can result in targeted advertising and behavioral modifications. In the hands of the state, it can result in immense and all-pervasive repression. Liberty was absent in 1984 due to an oppressive state with limited visibility, liberty is absent for Uyghur populations as a result of a digital panopticon fed by data, linked by networks, and rule by AI.