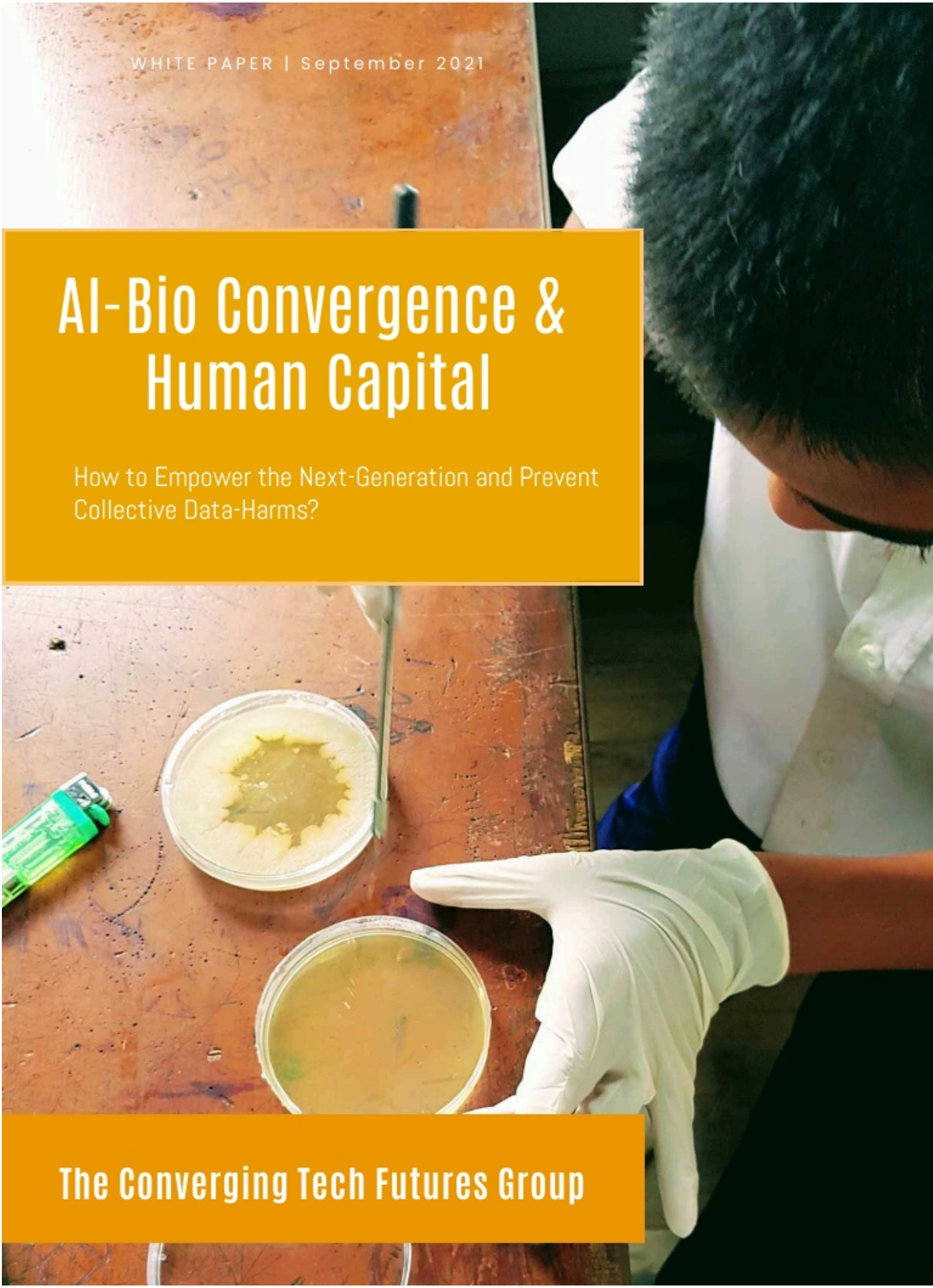


WHITE PAPER | September 2021

AI-Bio Convergence & Human Capital

How to Empower the Next-Generation and Prevent Collective Data-Harms?

The Converging Tech Futures Group

AI-BIO CONVERGENCE & HUMAN CAPITAL

How to Empower the Next-Generation &

Prevent Collective Data-Harms?

By Eleonore Pauwels

White Paper | September 2021

The Converging Tech Futures Group

Table of Contents

PARADIGM SHIFT

Synergistic combination of multiple technologies

Driven by digital infrastructure and AI

- Speed of discovery/ application
- Precision
- Transformative scale
- Personalization

POWERED BY AI

UNDER PINNED BY

Ubiquitous Connectivity | **Big Data** | **High Speed Computing**

COVID-19 Vaccines

- **Bio:** Rapid genomic sequencing of virus
- **Nano:** Lipid nanoparticles deliver strands of m-RNA without being degraded by body's defences
- **Info:** Used at all stages

Climate-Smart Agriculture

- **Bio:** Synthetic biology to enhance photosynthesis or absorption of nutrients
- **Nano:** Embedded plant sensors that communicate with electronic devices
- **Info:** Crop and water sensors for real-time moisture measurement and irrigation; Machine-to-machine interaction; Robots, drones

Adaptive Learning

- **Cogno:** Link between emotion and learning
- **Info:** sensors for capturing facial expressions and body parameters
- AI based "affective computing" to understand learning style and suggest changes

The Author

Eleonore Pauwels is an international expert in the security, societal and governance implications generated by the convergence of artificial intelligence with other dual-use technologies, including cybersecurity, genomics and genome-editing. Pauwels provides expertise to the World Bank, the United Nations, the Konrad Adenauer Stiftung Foundation and the Global Center on Cooperative Security in New York. She also works closely with governments and private sector actors on AI-Cyberthreats Prevention, the changing nature of conflict, foresight and global security.

In 2018 and 2019, Pauwels served as Research Fellow on Emerging Cybertechnologies for the United Nations University's Centre for Policy Research. At the Woodrow Wilson International Center for Scholars, she spent ten years within the Science and Technology Innovation Program, leading the Anticipatory Intelligence Lab. She is also part of the Scientific Committee of the International Association for Responsible Research and Innovation in Genome-Editing (ARRIGE). She writes for Nature, The New York Times, The Guardian, Scientific American, Le Monde, Slate, UN News, The UN Chronicle and The World Economic Forum.

AI & CONVERGING TECHNOLOGIES FOR HUMAN CAPITAL

EXECUTIVE SUMMARY: We are entering an era of hybrid opportunities and threats generated by the combination of artificial intelligence (AI) and other powerful dual-use technologies, with implications for nearly every aspect of daily lives. The convergence of AI and affective computing, cyber and biotechnologies, robotics and additive manufacturing raises complex global implications that are poorly understood, leaving the multilateral system with limited tools to anticipate and prevent emerging risks. At the same time, the spread of AI convergence across a wide range of States, non-State and transnational actors and entities means that the challenges of tomorrow must be addressed collectively and innovatively. This paper aims to assess the potential of converging technologies to accelerate human capital outcomes, including by promoting societal inclusion and empowering individuals at the household and community levels. While the paper introduces important positive shifts from technological convergence, its main arguments provide an in-depth analysis of the potential harms to populations generated by dual-use technologies.

□ To what extent are converging technologies a paradigm shift for human capital, drastically transforming how knowledge is produced and evaluated, altering what type of learning is valued and made available to youth? As explained in this paper, we face new knowledge and power asymmetries.

□ What are promising areas where clusters of technologies could provide either quick wins or drastic interventions to improve human capital outcomes? What are the short-term and medium-term risks to human capital outcomes generated by the convergence of powerful technologies?

□ What are the ethical and legal implications that the international community should consider when integrating converging technologies to its human capital projects? What safeguards at the operational and project levels are needed to promote accountability while harnessing converging technologies for human capital development?

□ In the longer-term, is there a role for the multilateral system to conceive pathways that would enable citizens and researchers within democratized innovation ecosystems (citizen science, fabrication labs, frugal innovation) to harness their own designs and datasets?

STRUCTURE OF REPORT: This paper starts with an introductory analysis of the deep gender-based and socio-economic inequalities in the South Asia region, highlighting specific technological risks that may drastically deepen these inequalities. Section 1 also provides a risk stratification that multilateral organisations could use for risk prevention and mitigation. Section 2 explains the paradigm shift that converging technologies represent for human capital outcomes, defining concepts and providing a matrix of technological functions that will serve human capital outcomes across systems. Section 3 presents several case-studies (education; health/nutrition; democratized innovation) that elicit the dual-use potential of converging technologies for human capital, focusing on opportunities but also specific risks to civilian and vulnerable populations. Section 4 analyses crucial ethical considerations for the multilateral system and international development community to ponder when planning to harness converging technologies for human capital. Section 5 concludes with strategic and operational recommendations for the multilateral system and international development community when it comes to converging technologies, data-protection and human capital outcomes.

TAKE-AWAY MESSAGES

□ AI is a transformative paradigm for human capital because it is essentially replacing the existing epistemic methods developed by humans and societies to produce knowledge and assess its value. Knowledge-production is increasingly automated by algorithms away from our explanatory scrutiny.

□ The convergence of AI and powerful data-capture technologies is giving rise to "affective computing," algorithms that can analyse us, predict our behaviours and emotions with drastic impact for education and employment. Youth and adults will have to adapt to an era, not only of automation, but also of competition with algorithms for cognitive and creative performance. Population subgroups could be excluded from economic flourishing, owing to both, lack of jobs and relevant skills.

□ The global private sector is invested in running and owning strategic elements of critical public infrastructure in health (hospitals, medical insurance companies) and education (schools, Ed Tech). There is a risk to lead to a "private automation and optimization" of human capital.

□ The global supply chains which produced converging technologies are complex and fragmented. Corporate self-regulation is not sufficient. We face a significant accountability gap and pervasive misalignment between ethics and business incentives.

□ Converging technologies exhibit functions that can be misused. In particular, digital interdependence in tech convergence will lead to heightened cybersecurity risks with implications for civilian datasets and critical information infrastructure. Security and societal implications of these technologies need to be anticipated and assessed before deployment in vulnerable context (underserved minority youth, deeply unequal societies). We need to develop a "theory of No-Harm" and practice "socio-technical system" analysis.

□ Automated behavioural and emotional analysis is already used by digital platforms, data-brokers and intelligence corporations for citizens' profiling and micro-targeting, pointing to the risk of social scoring. This prospect concerns not only individuals, but also populations (crowd behavioural analysis). It is difficult to isolate converging technologies from their potential for surveillance by corporations, state and non-state actors. Regulatory frailty is a pervasive concern for human rights infringements.

□ The use of AI-led behavioural and emotional data-analysis in education has corrosive implications for child rights. Legal reflections should centre on issues of proportionality, data-purpose and data-minimization. Weak scientific foundations and predictive value of AI emotional analysis in learning should also be addressed. Finally, the issues of human flourishing, social development and child educational benefits should be recognized as they are in the UN Convention on the Rights of the Child. The Convention clearly states the need to act in the child's best interests (Art. 3), the child's right to freedom of thought (Art. 14) and privacy (Art. 16), the right to develop full potential (§1 Art. 29), the child's right to liberty (§2 Art. 29), and the child's right to be protected from economic exploitation (Art. 32).

□ Yet, another vision exists for human capital where converging technologies are designed for serving empowerment, gender inclusion and civic participation. The real strength of democratized innovation ecosystems goes beyond the technical; learning is conceived as a social experience based on in-generation and inter-ethnic/economic solidarity. What drives innovation in these spaces involves a common ethos characterized by de facto interdisciplinarity, open-source and peer-to-peer knowledge-sharing, increased self-esteem, acceptance and empathy. This is why they provide an innovative and alternative incubator for individual and collective empowerment using cross-discipline synergies and serving local needs. Mentorship help students become local problem-solvers from cradle to career. Democratized innovation ecosystems provide a collaborative education environment where students learn how to use converging technologies (computation, engineering, and biotechnologies) to become entrepreneurs, creative agents of change, and innovators in the classroom and their communities. Moreover, this form of empowerment is not only open to students but also to STEM field teachers, providing them with a sustainable exposure to emerging technologies that can contribute to more formal classroom teaching.

SECTION 1 – NEW POWER ASYMMETRIES AND TECHNOLOGICAL IMPLICATIONS IN AN INCREASINGLY UNEQUAL WORLD

Politically, legally and ethically, our societies are not properly prepared for the deployment of AI and converging technologies. At national and international levels, we lack a comprehensive understanding of the threats that AI and converging technologies can pose at the individual human level, broader threats to populations, and geopolitical confrontations potentially triggered by the combination of new technological trends.

In the future, governments in the South Asia region (SAR) will need to develop a common understanding of technological convergence to be able to design proper oversight in collaboration with strategic actors in the private sector and civil society. States lagging behind in AI and converging technologies are the most at risk and the least likely to have any foresight capacity and this is the case for most SAR countries at the exception of India.

Pressured in a race to develop AI talents, research and industrialization pathways, governments are not equal in their ability to understand and anticipate evolving security risks, including in their political and socio-economic dimensions. Most national AI strategies have been developed by tech-leading states, leaving large unprepared regions in this new geopolitical landscape of converging technologies.

Most AI national strategies are treating issues related to civilian/human security, inequality and socio-economic cohesion as a secondary policy priority behind R&D and, more importantly, do not provide concrete mechanisms (audit, standards, redress system) to ensure such policy priorities go beyond aspirations.[1] Governments are struggling to foresee the broader, transformative landscape of AI and converging technologies and its converging security threats to civilian populations and critical information infrastructures. To avoid deepening tech-driven exclusion, an urgent priority may be to create new incentive structures and modes of public-private sector collaborations that imperatively integrate mechanisms of transparency and meaningful accountability. Such structures could create incentives and guidelines to ensure more founders in the private sector take on some of the real-world problems faced by SAR countries and do so with ethics, transparency and accountability at their core.

Another imperative is to create, in SAR countries, opportunities for academics and entrepreneurs to develop channels of learning and collaboration for the next generation to gain research, engineering and applied skills in AI and converging technologies. There are also opportunities for SAR countries with respect to AI and converging technologies. Data-optimization and predictive intelligence, in combination with biotechnologies, will enable innovation in sectors as vast as climate change, precision health and agriculture, personalized education, and additive manufacturing, all activities that might be more transformative for the SAR region. The capacity of AI to fight corruption and improve government efficiency, service delivery and public administration is too often forgotten. Cooperation between SAR countries, the World Bank and international financial or development institutions could harness predictive intelligence for public good. This, however, comes with a significant caveat: the private sector leading AI and technological convergence would have to help SAR countries avoid the array of hybrid threats described in the sections below, and facilitate understanding of how complex algorithms operate and impact already vulnerable societies. The ultimate goal of integrating AI and converging technologies in human development strategies would be to promote endogenous responsible innovation and strengthen economic and social resilience.

[1] AI Now Institute, 2019. ''AI Now 2019 Report''. https://ainowinstitute.org/AI_Now_2019_Report.pdf

The covid-19 pandemics is not only creating new corrosive inequalities, in particular for women and girls, but it also magnifies existing patterns of poverty and disempowerment. In a matter of mere months, the coronavirus has annihilated global gains that took two decades to achieve, leaving an estimated 70 to 100 million people at risk of extreme poverty.[2] In the long-term, the pandemic will keep impacting vulnerable communities with drastic human capital erosion, in particular for those already on the margins of educational systems, with more socio-economic exclusion, and potentially lethal sustained outcomes for children in their early years.

RISING GENDER-BASED INEQUALITIES

The current situation and future prospects are particularly dire for women and girls in the South Asia region, which will most likely experience its worst economic performance in 40 years. A recent report from the UN explains that the pandemic is not only magnifying gender inequalities, but even threatening to wipe out decades of progress in the work place.[3] The International Labour Organization foresees that the global slowdown could have daunting socio-economic implications for women in South Asia, where a substantial amount of working women are employed in the informal and care economy with minimal, or inexistent, protections.[4] The economy of care and domestic duties drastically limit the time women can search for jobs. In India, women perform 9.6 times more unpaid care work than men, about three times the global average. The pandemic has increased that burden for many women, according to the International Labour Organization.

Across every sphere of human capital, from health to the economy, education and employment to social protection, the impacts of COVID-19 are exacerbated for women and girls simply by virtue of their sex. Compounded economic impacts are felt especially by women and girls who are generally earning less, saving less, and holding insecure jobs or living close to poverty. As a result of the impact of COVID-19 on medical infrastructure, the health of women generally is adversely impacted through the reallocation of resources and priorities, including sexual and reproductive health services.

Such devastating setback raises crucial questions about how to further protect human capital from erosion and what potential converging technologies could bring to that mandate. If designed with accountability, privacy and local empowerment at the core, converging technologies bear significant promises in optimizing access to and delivery of services essential to human capital. But such values have to be imperatively built into technological design, deployment and governance, which is often not achieved (see Section 3 & 4).

A sector where converging technologies have exceptional potential is democratized health. Neural nets are already used in India to diagnose retinopathy and offset the shortage of doctors in rural neighbourhoods.[5] Equipped with AI analytics software, sensors, and cameras, mobile phones can become a diagnostic tool, increasingly used for digital microscopy, cytometry, immunoassay tests, and vital signs' monitoring. For instance, in a mobile phone equipped with a camera, image recognition systems can detect malaria in a blood smear, or screen for pap smear and cervix cancer remotely.

The potential of democratized innovation with converging technologies includes epidemics and crops monitoring as well as medical equipment delivery. Portable genomics sequencers bring the lab to the jungle, allowing for the diagnosis of Ebola viruses (and covid-19 virus) in real-time. In Shenzhen's Open Innovation Lab, young inventors have designed wearable devices that rely on image recognition to

[2] https://blogs.worldbank.org/opendata/updated-estimates-impact-covid-19-global-poverty

[3] https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2020/policy-brief-the-impact-of-covid-19-on-women-en.pdf?la=en&vs=1406

[4] Idem

[5] See: Raju M., et al. « Development of a Deep Learning Algorithm for Automatic Diagnosis of Diabetic Retinopathy.'' Studies in Health Technology Informatics; 245: pp. 559-563. http://ebooks.iospress.nl/publication/48210. Also see: Comstock J. 2016. "Google researchers use deep learning to detect diabetic retinopathy with upwards of 90 percent accuracy." Mobile Health News; 29 November. https://www.mobihealthnews.com/content/google-researchers-use-deep-learning-detect-diabeticretinopathy-upwards-90-percent-accuracy

help farmers detect diseases on crops. Companies like Zipline are using AI technology in autonomous drones to deliver critical medical supplies, such as vaccines, to rural hospitals in Africa.

A simple but life-changing breakthrough is the use of AI to optimize the delivery of services by community health workers in underserved rural areas. For instance, in Bangladesh, community health workers can use AI to know where and when to be present for birth deliveries, which contributed to a boost in neo-natal survival rates by over 30 per cent.[6] In a similar vein, community health workers in Tanzania now deliver targeted health campaigns to pregnant women and young caregivers using an AI-powered mobile application.[7] Meanwhile, researchers in Brazil rely on machine learning algorithms to forecast the need to resuscitate new-borns suffering from birth asphyxia, a condition still endemic in developing countries.[8]

Using AI to democratize health and education (see Section 4, case study 1 & 2) is powerful, yet it raises important questions of data-ownership, privacy and security, digital empowerment and self-determination/agency. Questions also abound about the safety of algorithmic design and the hidden biases that could discriminate students and corrupt health decision-making if training datasets do not carefully mirror the health and cognitive determinants of local populations.[9] Structural inequalities and vulnerabilities could also be on the rise if AI health and education services are mainly provided by private companies, which do not allow for transfer of technology, skills, algorithms and data to local educators, doctors and hospitals. Helping raise these concerns should be part of the theory of change and a "Theory of No-Harm" in human capital.

PRIVATE SECTOR' S BEHAVIOURAL SURVEILLANCE FOR PROFIT

Two of our case-studies (1 & 2 in Section 4) explain in depth and details what key ethical issues and human right implications could lead to limit agency and empowerment for SAR populations, in particular for children and the next-generation of students. As we explain through our case-studies and ethical analysis, converging technologies and increasing data commodification is already creating new knowledge and power asymmetries (see Section 3).

The World Bank recently came across an example of how practices of children's data monetization in Ed Tech can deepen the accountability gap, but also impoverish the same vulnerable families in need of protection and support. The Bridge International Academies (BIA) program has led to recent controversy over its plan to monetize the data it collects about school children through their education career. Since its inception, the BIA program has scaled to operate in about 520 (pre)primary schools in Sub-Saharan Africa and is now planning to run about 4000 schools in the Indian state of Andhra Pradesh. Questions abound about the scientific quality of its curricula and the adequacy of BIA's teaching methods: the role of teachers is being reduced to handling a very strictly scripted curriculum, designed by far-away innovators, and automated on a tablet which collect students' behavioural data.[10]

Concerns over children's behavioural data monitoring are even more sobering and call for urgent, in-depth ethical and legal reflections in Ed Tech. Under the Bridge programs, children's data collection does not stop at emotions but also include economic status, financial and transactional data of each family involved. As explained by Pawelec, the BIA program acquires data related to the "payment history for every pupil […] 3 million monthly transactions including timeliness and completeness, [i.e.]

[6] Brunskill E. and Lesh D. 2010. "Routing for Rural Health: Optimizing Community Health Worker Visit Schedules." AAAI. http:// www.aaai.org/ocs/index.php/SSS/SSS10/paper/download/1139/1369

[7] UNICEF. 2017. Annual Report, Tanzania. https://www.unicef.org/about/annualreport/files/Tanzania_2017_COAR.pdf

[8] Reis MAM., et al. 2004 "Fuzzy expert system in the prediction of neonatal resuscitation." Brazilian Journal of Medical and Biological Research; 37(5): pp. 755-764. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-879X2004000500018 &lng=en&tlng=en

[9] https://ainowinstitute.org/AI_Now_2019_Report.pdf

[10] http://globalinitiative-escr.org/wp-content/uploads/2017/07/July-2017-10-lessons-learned-from-two-recent-media-articles-on-Bridge-International-Academies.pdf

enough data and insight to create a credit-scoring system, [and] 50 million pupil attendance records."[11] Such insights are invaluable for BIA, which has started plans to "sell the data for commercial purposes to companies creating new financial services and low-cost health insurances." Such services are designed for those who *score* on less than $2 a day.

While BIA imposes extremely strict payments schedule on families, the program is able to mine populations' data (children and their families) by selling such information for commercial targeting and insurance profiling. This is troubling when investigations have shown that "fees and other costs charged at Bridge, combined with a strict enforcement system whereby children are expelled for missing payments, means that either children miss schools or families miss on other essential services such as healthcare. This may also largely explain the low and declining enrolment."[12]

The convergence of AI with other emerging technologies not only creates potential to widen inequalities, but may also increase security vulnerabilities for large swaths of civilian populations, including children and young adults.[13] Think of autonomous malware able to shape forms of social engineering and cyber-bullying tailored to manipulate the trust and lack of cybersecurity awareness of young children or other vulnerable users. The same algorithms can gradually learn from emotion analysis to manipulate the next generation of online students. Information disorders have been easier to scale and better at targeting anxious citizens during the current pandemic crisis.

STATE SURVEILLANCE FOR POPULATIONS' REPRESSION AND CONTROL

The ability of converging technologies to capture large amounts of sensitive behavioural, emotional and biological data about individuals and populations and subsequently control private human behaviour also has direct implications for agency and human capital. New forms of social, political and behavioural control are already stretching current approaches to monitor and implement human rights globally, and will certainly require to anticipate and better understand the societal implications of dual-use technologies. A recent investigation by the Australian Strategic Policy Institute demonstrates how police forces in China are collecting blood samples from men and boys from across the country to build a genetic database of its roughly 700 million males, as a powerful new way to subject citizens to high-tech surveillance.[14] The project is a major new step in China's efforts to use genetics to control its people, which had already been used for automated ethnic profiling of the Uighurs populations. The surveillance state also includes advanced cameras, facial recognition systems and artificial intelligence.

Sophisticated surveillance is increasingly imported to countries in Sub-Saharan Africa as well as South and East Asia.[15] In this regard, responses to the covid-19 crisis have accelerated large scale behavioural surveillance, legitimizing the need for personal and medical data collection, monitoring of populations' movements, and screening of mobile phones' activity and contacts. For instance, Vietnam has relied on massive social closures and extensive surveillance of citizens. From the early onset of the pandemic, Vietnam shuttered non-essential businesses and schools and enacted large-scale quarantines—tens of thousands of citizens have been placed in "quarantine camps" run by the military.[16] Vietnam's aggressive monitoring and surveillance of citizens has been supported by the government's large network of informants, which has helped to identify and quarantine those suspected of infection and those who have been in contact with them.

State surveillance also increasingly threatens to close virtual civic spaces. The Government of Thailand has declared new executive powers to combat "fake news" related to Covid-19 and the government's

[11] Pawalec M., 2018. "How Bridge International's Tech-Driven For-Profit Schools in Africa May Reinforce Global Power Inequalities". *Sichesrheits Politik Blog*. https://www.sicherheitspolitik-blog.de/2018/06/19/how-bridge-internationals-tech-driven-for-profit-schools-in-africa-may-reinforce-global-power-inequalities/#fnref-9033-22

[12] http://globalinitiative-escr.org/wp-content/uploads/2017/07/July-2017-10-lessons-learned-from-two-recent-media-articles-on-Bridge-International-Academies.pdf p 2

[13] https://www.weforum.org/agenda/2020/06/prevent-cyber-bio-security-threats-covid19-governance/

[14] https://www.nytimes.com/2020/06/17/world/asia/China-DNA-surveillance.html

[15] https://www.comparitech.com/blog/vpn-privacy/surveillance-states/

[16] https://www.csis.org/analysis/strengths-and-vulnerabilities-southeast-asias-response-covid-19-pandemic

response to the epidemic, including censoring media, social media, and personal communications.[17] In Thailand, the new emergency powers extend to arresting and detaining those charged with sharing false information, which critics fear will be used liberally against political opponents. The Myanmar government has also threatened to crack down on "fake news" reports in media and social media that do not align with official reporting on the pandemic.

But it is India that has recently and powerfully upgraded its technological surveillance capacities (see Section 3) to deploy individual facial recognition, algorithmic crowd analysis and mobile biometrics identification.[18] In Delhi, police officers are increasingly using facial recognition devices to screen individuals entering protest venues. In Chennai, surveillance drones have been used to perform behavioural crowd-analysis during protest marches. And in Hyderabad, there has been rising instances of law-enforcement collecting individuals' biometrics features on the spot to check for past criminal activity.

In recent years, law-enforcement authorities have started regularly collecting soft biometrics features, using facial recognition and mobile fingerprinting, on individuals deemed "suspicious" without evidence of criminal activities. Such policing and surveillance practices often target poor neighbourhoods heavily populated by Muslims and populations from north India. Capturing bulk, mass biometrics features of populations in public spaces is a form of pervasive surveillance that essentially treats every individual as a potential suspect, subject to continuous investigation. At present, there are no existing robust mechanisms to ensure that legislative oversight can prevent the proliferation of surveillance activities by India' security agencies.

Biometrics data collection and centralization for welfare and civil services had also significant implications for the most vulnerable among India's large population. In the last five years, the Indian government has mandated compliance with the creation of a country-wide biometrics database as part of Aadhaar's identification profile.[19] Registration and biometrics data collection for Aadhaar has been coercive because it is either de facto or legally mandatory to be enrolled to access essential welfare services. The logic behind such centralised system is that it will combat fraud by identifying fraudsters or "fake beneficiaries" of welfare services. Yet, in the context of Aadhaar, a growing number of welfare-benefits denials have penalized some of the most vulnerable individuals and families that were either not properly enrolled or could not authenticate due to technical failures. In these specific cases, failures led to higher levels of malnutrition and even starvation deaths.[20] Authorities could increasingly use Aadhaar as a way to unilaterally exclude minorities from social protection – a concern raised before the Supreme Court in the 2018 hearings, but with no redress or legal response.

Moreover, the Aadhaar biometrics database has been criticized for its design that allowed commercial surveillance.[21] Until intervention by the Indian Supreme Court, 300 any private entity was allowed to use the state's biometric ID infrastructure for authentication, including banks, telecom companies, and

---

[17] Idem

[18] https://www.foreignaffairs.com/articles/india/2020-02-19/indias-growing-surveillance-state

[19] Roy S. 2018. "Aadhaar: India's Flawed Biometric Database." The Diplomat; 6 March. https://thediplomat.com/2018/03/aadhaarindias-flawed-biometric-database/. Also see the Aadhaar website: https://uidai.gov.in/your-aadhaar/about-aadhaar.html

[20] Aadhaar Linked To Half The Reported Starvation Deaths Since 2015, Say Researchers', Huffington Post India, 26 September 2018 available at https://www.huffingtonpost.in/2018/09/25/aadhaar-linked-to-half-the-reportedstarvation-deaths-since-2015-say-researchers_a_23539768/

[21] See Aria Thaker, "The New Oil: Aadhaar's Mixing of Public Risk and Private Profit," Caravan, April 30, 2018, https://caravanmagazine.in/reportage/aadhaar-mixing-public-risk-private-profit; Usha Ramanathan, "Who Owns the UID Database?," MediaNama, May 6, 2013, https://www.medianama.com/2013/05/223-who-owns-the-uid-database-usha-ramanathan/ ; and Pam Dixon, "A Failure to 'Do No Harm'—India's Aadhaar Biometric ID Program and Its Inability to Protect Privacy in Relation to Measures in Europe and the U.S.," Health and Technology 7, no. 6 (June 2017), https://doi.org/10.1007/s12553-017-0202-6

a range of other private vendors with little scrutiny or privacy safeguards.[22] The Supreme Court did suppress a legal provision that gave private corporations access to the Aadhaar database, but, a few months later, the Parliament re-enacted that provision renewing similar access for the private sector.[23]

Authentication records created by biometrics ID systems, as well as the ability to aggregate information across databases, can also increase the power of surveillance infrastructures and discriminatory profiling available to governments. A recent investigation based on official documents leaked to the Huffington Post India shows that the Indian government plans to connect Aadhaar with a program called the National Social Registry (NSR) or Social Registry Information System.[24] Such integrated database could update in real-time personal and precise data about populations, including employment, marital status, financial transactions, and relocations to another state. In the absence of robust and accountable mechanisms to implement meaningful data-protection, Indian authorities' growing ability to track its citizens has the potential to be transformed into a social credit system, which may once again have discriminatory implications for the most vulnerable.

Source: Photo by Yogesh Rahamatkar on Unsplash

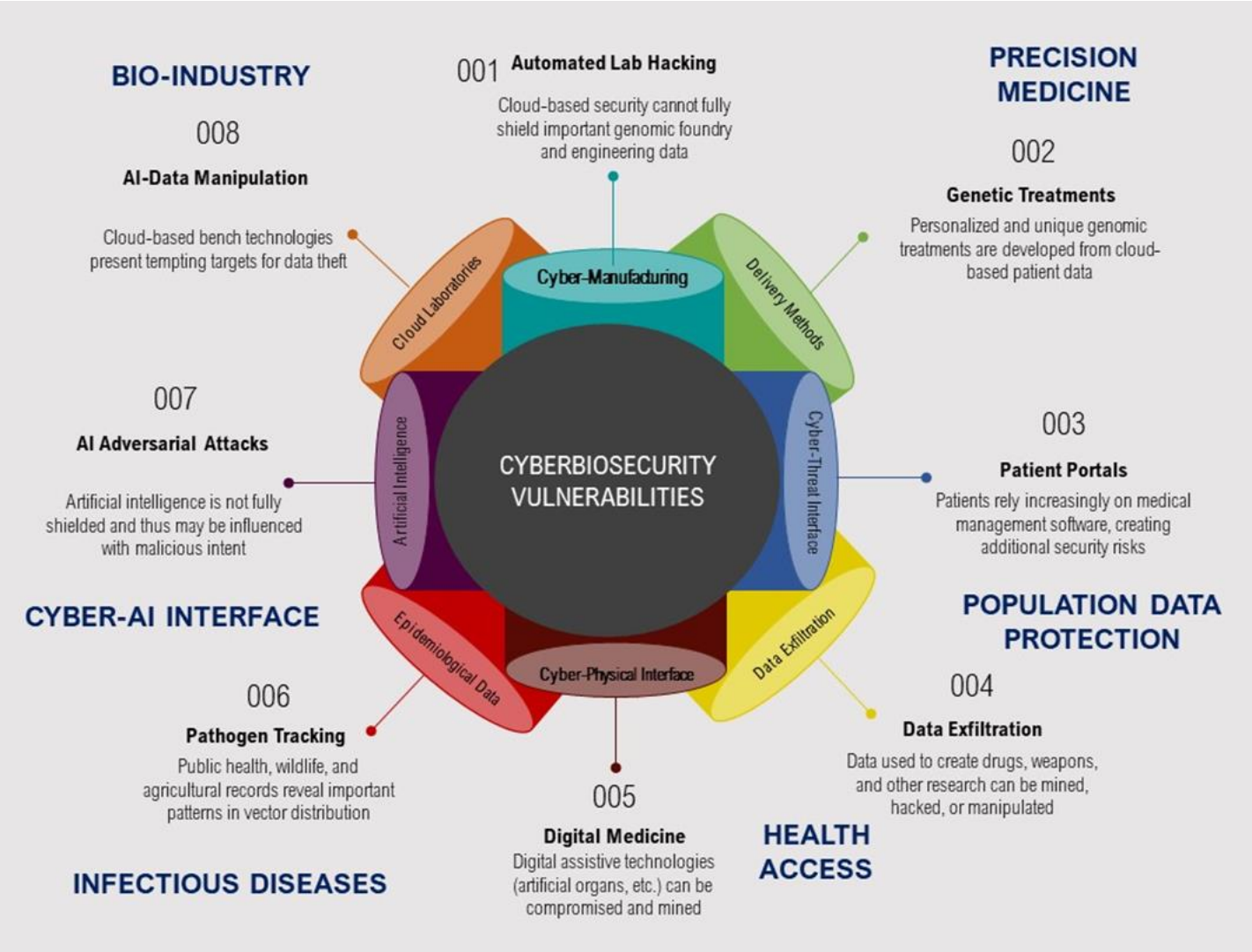
While surveillance technology has proliferated in India over the last decade, institutional and legal safeguards have not kept pace.[25] The Indian Parliament has yet to enact a data protection law, and the courts have failed to adequately grapple with the ethical and constitutional challenges posed by invasive new technologies. The Indian public, for its part, is not necessarily informed and empowered enough to understand the scale of state surveillance and advocate for meaningful debate, consent and redress processes in case of harm.

The below table offers a potential "risks' stratification" to help the Bank navigate future ethical, legal and accountability challenges in harnessing converging technologies for human capital.

[22] Vindu Goel, "India's Top Court Limits Sweep of Biometric ID Program," New York Times, September 26, 2018, https://www.nytimes.com/2018/09/26/technology/india-id-aadhaar-supreme-court.html
[23] https://thewire.in/law/aadhaar-rbi-supreme-court-uidai
[24] https://www.huffingtonpost.in/entry/aadhaar-national-social-registry-database-modi_in_5e6f4d3cc5b6dda30fcd3462
[25] https://www.foreignaffairs.com/articles/india/2020-02-19/indias-growing-surveillance-state

TIMELINE	RISK STRATIFICATION
Current	**DATA COMMODIFICATION FOR PRIVATE SECTOR PROFIT** -commodification of behavioural, emotional and biometrics data of children and other vulnerable populations for education scoring, future commercial targeting, and exclusion/discrimination schemes **DATA COMMODIFICATION & MANIPULATION FOR STATE SURVEILLANCE** -commodification of behavioural, transactional, socio-economic and consumption data for social credit systems, and exclusion/discrimination schemes -use of personal data to silence civil society resistance, repress traditional media structures and harm the reputation of knowledge-institutions, leading to closure of virtual civic space impacting population's resilience and social fabrics **INFORMATION DISORDERS, DISINFORMATION & HATE SPEECH** -use of personal, demographic, ethnic, behavioural and emotional data collected on children and populations for targeting disinformation and polarization, emotion manipulation and hate speech, and for radicalization -Mobilization of larger population subgroups around violent narratives including around elections **FAILURE OF TECHNOLOGICAL DESIGN AND PREDICTIVE VALUE** -biases in data-sets and algorithmic design as well as poor performance in predictive value may lead to system (access/delivery/optimization) failures, with corrosive implications for underserved groups
Now to next 5 years	**CYBER-OPERATIONS / CYBER-BULLYING / SOCIAL ENGINEERING** -Use of personal and emotional data for social engineering leading to more efficient, powerful acts of cybercrime -Use of biometrics data for precision biometrics attacks (cyberattacks where autonomous malware uses soft facial/voice/biometrics features for impersonation) -Exfiltration of sensitive datasets about populations to target attacks towards vulnerable subgroups (groups on the bridge of food insecurity; biometrics data used for retaliation against specific minorities; data about populations' movements for targeted attacks in conflict) - Automated Data-Poisoning: Poisoning data in critical information infrastructure, for instance related to medical/hospital databases, as well as biometrics, civic & electoral registries -Cyberattacks targeting automated supply chains with impact for food security and delivery of essential human capital services -Cyberattacks where autonomous malware weaponize other dual-use technologies (biotech, 3D printing, robotics, including drone technologies)

CRITICAL CIVILIAN TARGETS OF CYBER-BIOSECURITY THREATS[26]



[26] Source: https://www.frontiersin.org/files/Articles/451363/fbioe-07-00099-HTML/image_m/fbioe-07-00099-g001.jpg https://www.frontiersin.org/articles/10.3389/fbioe.2019.00099/full

SECTION 2 - CONVERGING TECHNOLOGIES: A PARADIGM SHIFT FOR HUMAN CAPITAL?

Global public health crises create a context of extreme fragility where converging technologies can be harnessed to improve large-scale crisis prevention and accelerate human capital outcomes, but also amplify human insecurity. We are entering an era of hybrid opportunities and threats generated by the combination of artificial intelligence (AI) and other powerful dual-use technologies, with implications for nearly every aspect of daily lives. The convergence of AI and affective computing, mobile biometrics and facial-recognition, cyber and biotechnologies, robotics and additive manufacturing, raises complex global implications that are difficult to understand and anticipate. This era of technological convergence leaves governments, particularly in context of crisis and fragility, with substantial challenges: How to harness technological opportunities for improving human capital, social resilience and empowerment? How to mitigate the potential for technologies to increase systemic vulnerabilities and distributive inequalities?

AN EPISTEMIC & LEARNING REVOLUTION AS MUCH AS A TECHNOLOGICAL ONE

● AI is a transformative paradigm for human capital because it is essentially replacing the existing epistemic methods developed by humans and societies to produce knowledge and assess its value. Knowledge-production is increasingly automated by algorithms away from our explanatory scrutiny.

● The convergence of AI and powerful data-capture technologies is giving rise to "affective computing," algorithms that can analyse us, predict our behaviours and emotions with drastic impact for education and employment. Population subgroups could be excluded from economic flourishing, owing to both, lack of jobs and a lack of relevant education and mental/emotional flexibility.

● What skills and forms of learning will be valued in children's education when didactic and pedagogic methods and values are transferred to online platforms deploying real-time behavioural analytics? What reskilling/retraining programs will be valued and deployed for adults who have to compete in cognitive and creative skills with algorithms?

●This revolution calls for in-depth 1) epistemic reflections on what we value in education and knowledge production (e.g. the role of predictive analytics versus diverse forms of knowing), 2) ethical reflections about inclusion and empowerment ("knowing together," mentorship, peer-to-peer experience), and 3) legal reflections on knowledge and power asymmetries (ownership of data & algorithms involved in knowledge-production; who decides who learns what and how? For what kind of empowerment?).

AUTOMATION OF LEARNING & KNOWLEDGE PRODUCTION

Algorithms are becoming responsible for quantitative and qualitative analyses about populations and situations in an increasing range of sectors, from medicine, education, fraud detection in welfare, to the retail industry and the field of criminal justice. As algorithms can be trained to identify data patterns and interference in extremely large datasets, the computational methods these algorithms rely on to produce knowledge is used for quantitative analysis (populations' migration at a border, for instance) and qualitative analysis (diagnosing retinopathy specific to a single patient). In brief, the AI revolution automates knowledge-production far beyond "big data" analysis.

Let's take the medical field (diagnosis) as an example. Modern medicine faces a new array of temptations and opportunities to essentially transform a field in which contextual expertise, human intelligence, trust and interpersonal skills represent significant attributes into a powerful science of predictive and automated behavioural, physiological and biological analysis.

For instance, using their exceptional capacity for image pattern recognition, algorithms are already surpassing doctors' eyes in distinguishing benign moles from melanomas,[27] or small brain bleed from lethal embolism.[28] Deep learning algorithms are used for reading X-rays, CT scans and MRIs of every variety. Soon, they will be used for pathological diagnoses, analysing Pap smears, listening to heart sounds, or predicting relapses in psychiatric patients. They are becoming excellent at knowing "What" (diagnosis) and "How" (few pixels make a difference between benign and malign moles). While newly practicing doctors have to be taught and trained (human learning curve), algorithms keep learning on their vast training datasets and keep updating each other if connected. These potential advantages of connectivity and "updatability" are so significant that, at least in some lines of work (and from a rentability perspective), it might make sense to replace most humans with computers, even if individually some humans still do a better job.

THE BLACK BOX – AWAY FROM OUR CRITICAL SCRUTINY

Yet, here lies an important epistemic and cognitive trap, what is commonly called the "black box." Algorithms have no explanatory capacity: they don't know and cannot explain "Why" a brain bleed is turning into an embolism and we don't know how and why they achieve such decision-making process. All the internal computational processing that allows algorithms to learn happen away from our scrutiny. Now, a doctor would rightly point out that diagnosticians and radiologists are not merely engaged in a binary, yes-no classification. They are not just detecting the embolism that led to a stroke. They are noticing the small bleed elsewhere that might make it lethal to use blood-thinner drugs; they are picking up on a nearby still asymptomatic but cancerous tumour. Doctors also search and ask "Why" and "Why now" in a personal human story or in a broader population trend. Which begs an important question: What is gained and lost in automating predictive AI analysis?

Such epistemic and cognitive revolution, in which knowledge-production is automated away from our explanatory scrutiny, has significant implications for the field of human capital:

● **BIASES, FAIRNESS & EMPOWERMENT**: First, this form of automated/algorithmic diagnosis could magnify risks that are difficult to anticipate such as false positives (diagnosing something that does not exist) or biases, and could optimize situations in ways that may conflict with our societal values (refusing medical care or welfare in situations that are not optimal by algorithmic rules). Algorithms are often "imported" from the North to the South, or from the designers and experts to the "underserved." The author of the Coded Gaze, Joy Buolamwini from MIT Media Lab, explains how current facial recognition algorithms, in their functioning nature and optimization processes, fail to discern the features of African-Americans, particularly women.[29] The new field of emotion-analysis struggle to detect the "smiles" of Asian women as they present in subtle modes, somehow different from original datasets.

What do fairness and empowerment mean in a world where only a small proportion – about 0.004% of the global population – have the knowledge and power to build machines that are intelligent enough to potentially decide who wins on the global job market, who can obtain insurance, or whose DNA or behavioural patterns will be mined by far-away marketers? Never have we faced a technology like AI – whose design is in the hands of a few, and who are mostly born in societies of abundance, yet a technology powerful enough to shape and impact the lives of vulnerable populations in the global south. This asymmetry of knowledge and power raises significant challenges for human capital.

---

[27] Esteva A., Kuprel B., Novoa R., Ko J., Swetter S., Blau H., and Thrun S., 2017. "Dermatologist-level classification of skin cancer with deep neural networks". *Nature.* https://www.nature.com/articles/nature21056

[28] Keane P. and Topol E., 2018. "With an eye to AI and autonomous diagnosis". *npj Digital Medicine* https://www.nature.com/articles/s41746-018-0048-y

[29] Buolamwini J. and Gebru T., 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". *Proceedings of Machine Learning Research.* http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Buolamwini J., 2018. "When the Robot Doesn't See Dark Skin". *The New York Times.* https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html

● **AI DATAFICATION & REDUCTIONISM IN EDUCATION**: Second, the strengths and limits of algorithmic analysis raise important questions for education. Will algorithms detect and stigmatize all types of learning disabilities without an explanation (or even a simple intelligent observation) of the root-causes? Will they induce a new form of reductionism, narrowing down holistic education, rewarding only certain type of knowledge? How will it impact the mentorship relation (teacher-student) in local communities? Will it frame education goals mainly towards market or product-driven logic? Will we witness a constant competition between humans and machines for cognitive and creative performance? What is the impact on human capital?

The combined analysis of soft biometric features, behavioural and physiological data is also giving rise to "affective computing" – algorithms that are trained to analyse, nudge, and communicate with us. This form of emotional analysis is supposed to improve human-machine interactions in applications that support precision medicine and targeted education. For instance, Affectiva, one of the leading affective computing companies, is interested in quantifying emotions to "get a complete understanding of individuals' overall wellbeing" as a suicide prevention tool.[30] Affectiva is also interested in using emotional and behavioural engineering to support children with autism and other learning disabilities.[31] Yet, beyond these aspirations, emotion analysis is already used by powerful technological platforms and data-brokers to collect and analyse massive amounts of personal and behavioural data about populations to predict and engineer future needs, demands and create lucrative markets.[32] Affective computing may incidentally lead to AI systems that may outperform human doctors, medical field workers, drivers, soldiers, bankers and teachers. This prospect has implications for human capital, calling for lifelong learning and, even more, a revolution in education.

● **CHILDREN'S EMOTIONAL & SELF-AGENCY:** The growing capacity of algorithms to target, influence, and even manipulate emotions, has been demonstrated in experiments performed by digital platforms. In 2012 and 2014, Facebook conducted its "emotional contagion" experiments when the company exploited social comparisons on its pages and relied on users' psychometrics profiles and subliminal cues to make people feel sadder or happier.[33] For Facebook, such experiments led to important findings: it is possible to manipulate online cues to influence the real-world behaviours and emotions of populations' subgroups; and this could be done in a subliminal way, without raising users' awareness and, subsequently, without users' consent. Intelligence collection based on emotional and behavioural data has also reached the toys' market. An example is the *My Friend Cayla* smart doll[34] that sends voice, thoughts, desires and other emotional data of the children who play with it back to the company's cloud. The product led to a US Federal Trade Commission complaint and a ban on the doll in Germany.[35]

For decades, the behavioural surveillance industry has been functioning in the shadow of our digital economy. Leading internet platforms have mastered converging technologies for shaping human behaviours towards commercial objectives. This new form of behavioural engineering suffers from

[30] For more information on Affectiva's SDK project for suicide prevention, see: Affectiva, 2017. "SDK on the Spot: Suicide Prevention Project with Emotion Recognition". https://blog.affectiva.com/sdk-on-the-spot-suicide-prevention-project-with-emotion-recognition

[31] For more information on Affectiva's Brain Power System project for autism, see: McManus A. 2017. "Brain Power Launches Anticipated Emotion-Enabled Autism System for Smart Glasses." Affectiva, 7 November. https://blog.affectiva.com/brainpower-launches-much-anticipated-emotion-enabled-autism-system-for-smart-glasse

[32] Zuboff S., 2020. "You Are Now Remotely Controlled". *The New York Times.* https://www.nytimes.com/2020/01/24/opinion/sunday/surveillance-capitalism.html; AI Multiple, 2020. "Affective Computing: In-Depth Guide to Emotion AI". https://research.aimultiple.com/affective-computing/

[33] Kramer A., Guillory J., and Hancock J., 2014. "Experimental evidence of massive-scale emotional contagion through social networks". *Proceedings of the National Academy of Sciences.* https://www.pnas.org/content/111/24/8788

[34] For more information about the My Friend Cayla Doll, see their website available at: https://www.myfriendcayla.com/meetcayla-c8hw

[35] Nienaber M. 2017. "Germany bans talking doll Cayla, citing security risk." Reuters; 17 February https://www.reuters.com/article/us-germany-cyber-dolls/germany-bans-talking-doll-cayla-citing-security-risk-idUSKBN15W20Q

weak self-regulation.[36] In this context of regulatory frailty, what will be the unintended consequences of subjecting the next-generation of students to increasingly refined forms of emotion-analysis? What are the implications of living with self-learning machines that acquire knowledge and engineer emotional responses using reasoning and values that we don't understand anymore? In this context, human cognition loses its personal character and value. Individuals turn into data and data become the only metrics.

There are currently no agreed and stress-tested methods to assess the ethical, security and human rights implications of delegating some strategic elements of behavioural-analysis and education to automated predictive technologies. Therefore, a substantial accountability gap exists when no methods and cross-sector collaborations have been conceived and deployed to anticipate unforeseen misuses and long-term impacts of AI and data-capture technologies on vulnerable populations.

WHAT IF… THE POWER OF KNOWING TOGETHER

Yet, another vision exists for human capital where converging technologies are designed for serving empowerment and civic participation. Democratised innovation communities are growing bigger, more ambitious and more networked. These communities are surfing wider societal forces, including a thirst for actionable data; the rise of connectedness and low-cost sensor technologies; and a push to improve the transparency and accessibility of science.[37]

For instance, in Dhaka, the community Mechamind[38] aims to support the vision of an open science society in Bangladesh. The team of mentors teaches underprivileged kids to work on local innovation problems and develop skillsets for future tech-based industries. The community lab has four collaborative streams, including democratized biology, AI and STEM, enabling technologies and prosthetics. Kazi Rhaman, the Lab Director, cares about mentorship; she is both, a biotech teacher/consultant, and a mental health first aider. We met in September 2016 with hundreds of community labs' directors from the U.S., Mexico, India and China, at the MIT Media Lab to discuss governance models for a "biotech without borders" movement.

In Baltimore (USA), the Digital Harbor Foundation[39] (DHF) provides an exceptional example of a community lab that, beyond structured educational programming, supports youth with access to meals and transportation. DHF does not only integrate more than a thousand underserved minority students a year, but it also provides professional development opportunities to build teachers/educators' capacity to teach technology skills and integrate tech learning, innovation, and creativity into their own learning environments. In 2015, DHF served 1,381 K-12 youth and 300 educators with direct-service technology and engineering programming ranging from 3Dprinting to human-centred design. 76% of DHF youth served live under Federal Poverty Line standards, while 73% are minorities and 50% are females.

During the covid-19 crisis, Taiwan provided a larger scale example of active public involvement in epidemic management.[40] Digital maps created by teams of entrepreneurs and civil society hacktivists showed where masks were available in Taiwan. But these maps also allowed citizens to participate in the management of masks as a scare common good. Citizens were able to reallocate rations of masks through donations to those who most needed them, which helped prevent the rise of a black market. The participatory success stimulated the government to provide computational resources and bandwidth required to allow a version of this service that could serve the whole population.

---

[36] AI Now Institute, 2019. ''AI Now 2019 Report''. https://ainowinstitute.org/AI_Now_2019_Report.pdf

[37] Wiggins A. and Wilbanks J., 2019. "The Rise of Citizen Science in Health and Biomedical Research". *The American Journal of Bioethics*. https://www.tandfonline.com/doi/full/10.1080/15265161.2019.1619859; Pauwels E., 2017. "The Internet of Living Things". *Scientific American.* https://blogs.scientificamerican.com/observations/the-internet-of-living-things/

[38] https://www.mechamind.io/

[39] https://www.digitalharbor.org/whatwedo/youth/

[40] Lanier J. and Weyl EG., 2020. "How Civic Technology Can Help Stop a Pandemic". *Foreign Affairs.* https://www.foreignaffairs.com/articles/asia/2020-03-20/how-civic-technology-can-help-stop-pandemic

These are just three examples of an upstart revolution where citizens are deciding not to wait around for adapting education and innovation to local needs. Only a diversity of knowledge and experience will help foster diligent technical design, anticipate ethical failures, and minimize the risks of unintended harms. If we, as global citizens, want to reclaim our technological future, we need to carefully consider the risks involved and the sources of inequalities and disempowerment that hide in mundane algorithmic designs. We also need to unveil the incentives behind the business model that shapes algorithmic logic. Access to the knowledge and education required for designing and anticipating the role of an emerging technology like AI is still a luxury. How can our far-reaching algorithmic inventions be designed and governed so that they meet the ethical needs of a globalizing world?

The word "diagnosis," what AI is good at, comes from the Greek for "knowing apart." Deep learning will only become better at such knowing apart—at recognizing faces or distinguishing moles from melanomas. But knowing, in all its scientific, human and societal dimensions, goes beyond computation. And in the realm of education, perhaps the ultimate rewards come from knowing together.

FUNCTIONAL AUGMENTATION FROM TECH CONVERGENCE

● Automated behavioral and emotional analysis is largely used by digital platforms, data-brokers and intelligence corporations for citizens' profiling and micro-targeting, pointing to the risk of social scoring. This prospect concerns not only individuals, but also populations (crowd behavioral analysis). It is difficult to isolate converging technologies from their potential for surveillance by corporations, state and non-state actors. Regulatory frailty is a pervasive concern.

Technologies are becoming complex hybrid systems that are merging and enabling each other, with drastic changes in velocity, scope and system-wide impact.[41] The age of technological convergence builds upon the digital revolution with a global surge in computing power and big data capacities. At the core of this transformation, AI is a general-purpose technology[42] that can learn to program and automate the operations of other digital, physical and biological technologies. Essentially, the advent of AI consists of two fundamental changes in organizational and technological operations:

- The analysis of large datasets is now often delegated to algorithms as they can be trained to identify data patterns and interference where previously this work was done by humans. Quantitative and even qualitative analysis based on large datasets is increasingly optimized and automated by using algorithmic programs.

- Algorithms are recommending and/or making decisions that previously were made solely by humans. The optimization of data-analysis by algorithmic programs allows these programs to perform and even automate reasoning, task distribution and decision-making. Relying on such capacities, AI programs can enable autonomy in other technologies and industrial platforms that are critical to civilian populations' survival and well-being.

Beyond the above two fundamental changes, AI provides several functions that may play a significant role in applications and innovations aimed at accelerating human capital. This functional augmentation could be playing an increasing role in enhancing the performance of digital technologies for education,

[41] Schwab K. 2016. "The Fourth Industrial Revolution: what it means, how to respond." World Economic Forum; 14 January. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

[42] A general-purpose technology is a term coined to describe a new method of producing and inventing that is important enough to have a protracted aggregate impact. General-purpose technologies are technologies that can affect an entire economy. GPTs have the potential to drastically alter societies through their impact on pre-existing economic and social structures.

health and social protection. Yet, while there is a potential for AI and data-capture technologies to produce innovation in personalized learning, smart ID and tele-medicine, numerous private sector applications primarily focus on predictive behavioral analysis for monitoring populations, for instance in the retail industry and in predictive policing.[43]

□ **Human demographic, biometrics and facial recognition**: Algorithms can identify biometrics data – from facial features, iris scans, fingerprints, hand and ear lobe geometry – by detecting and processing patterns and shapes specific to an individual, such as segmenting and indexing someone's iris scan. Algorithms are also acquiring the ability to analyze visual data that constitute demographic and personal information about people, such as gender, race, and age, and to collect information such as if they are wearing mask or what clothes they are wearing. IBM offers the ability to identify people by such factors as skin tone or whether they are bald.[44] Among academics, the challenge of automatically estimating a person's age based on their photographic appearance has attracted more and more researchers and has gotten more accurate.[45] Gait recognition — the ability to identify people based on how they walk — has been studied for years as a means of identifying people.[46] Gait can also reveal injury or certain medical conditions.[47]

Biometrics and facial-recognition technologies are being deployed at an accelerating pace in the Global South, including across Sub-Saharan Africa with over 30 nations in the process of registering their populations' biometrics into centralized national databases.[48] Citizens are required to use these new digital modes in order to participate in political and social life, such as voting in elections and accessing financial, health and education services. Yet, biometrics and facial-recognition technologies are also increasingly used for surveillance and predictive policing. In Zimbabwe, the Chinese company, CloudWalk, is using its 3D light facial software, to detect the features of local populations.[49] In February 2019, the French company, Gemalto, announced a smart-policing collaboration with police forces in Uganda to deploy portable biometric ID devices that use algorithms to confirm a match on the spot.[50] These are just two examples of how converging technologies are making populations' digital bodies and minds traceable in real-time.

□ **Human action recognition**: With progress in image recognition (computer vision), algorithms can learn to interpret and understand the visual world. With progress in language and speech/voice recognition (natural language processing), algorithms can be taught to understand the languages

---

[43] AI Now Institute, 2019.

[44] Riley T. and Russo S., 2016. "Driving value from body cameras". *IBM.* https://www.ibmbigdatahub.com/whitepaper/driving-value-body-cameras; Joseph G. and Lipp K., 2018. "IBM Used NYPD Surveillance Footage to Develop Technology that Lets Police Search by Skin Color". *The Intercept.* https://theintercept.com/2018/09/06/nypd-surveillance-camera-skin-tone-search/

[45] Huo Z., et al. 2016. "Deep Age Distribution Learning for Apparent Age Estimation". *CVPR.* https://www.cvfoundation.org/openaccess/content_cvpr_2016_workshops/w18/papers/Huo_Deep_Age_Distribution_CVPR_2016_paper.pdf; Uricar M., Timofte R., Rothe R., Matas J., and van Gool L., 2016. "Structured Output SVM Prediction of Apparent Age, Gender and Smile From Deep Features". *CVPR.* https://people.ee.ethz.ch/~timofter/publications/Uricar-CVPRW-2016.pdf

[46] A survey of work in this field is at Ke S., Thuc H., Lee Y., Hwang J., Yoo J., and Choi K., 2013. "A Review on Video-Based Human Activity Recognition." *Computers*. 2. 88-131. https://www.researchgate.net/publication/285197344_A_Review_on_VideoBased_Human_Activity_Recognition

[47] Sánchez-DelaCruz E., Acosta-Escalante F., Wister M., Hernández-Nolasco J., Pancardo P., and Méndez-Castillo J., 2014. Gait Recognition in the Classification of Neurodegenerative Diseases. https://www.researchgate.net/publication/283673082_Gait_Recognition_in_the_Classification_of_Neurodegenerative_Diseases

[48] Burt C., 2020. "New approaches to identity altering government operations to take the spotlight at ID4Africa 2020". Biometric Update. https://www.biometricupdate.com/202003/new-approaches-to-identity-altering-government-operations-to-take-the-spotlight-at-id4africa-2020

[49] Woodhams S., 2019. "How China Exports Repression to Africa". The Diplomat. https://thediplomat.com/2019/02/how-china-exports-repression-to-africa/

[50] Mayhew S., '' Ugandan police deploy Gemalto tech for rapid capture of suspects' biometric data''. Biometric Update. https://www.biometricupdate.com/201902/ugandan-police-deploy-gemalto-tech-for-rapid-capture-of-suspects-biometric-data

spoken by humans and determine the meaning of texts and arguments. Relying on the convergence of these different subfields, algorithms can learn to analyze and interpret human actions, from recognizing simple human actions such as walking and running towards recognition of complex realistic human activities involving multiple persons and objects.[51] AI engineers are aiming not just to train computers to understand human actions, but also to predict them. Prediction of human movement (pedestrians, cyclists, and drivers) is vital for autonomous vehicles, but researchers are also seeking to go beyond that. As research by a team of Korean engineers argues:

"In many real-world scenarios, the system is required to identify an intended activity of humans (e.g. criminals) before they fully execute the activity. For example, in a surveillance scenario, recognizing the fact that certain objects are missing after they have been stolen may not be meaningful. The system could be more useful if it is able to prevent the theft and catch the thieves by predicting the ongoing stealing activity as early as possible based on live video observations."[52]

□ **Anomaly detection**: Algorithms' ability to identify patterns and anomalies in large data-sets makes it possible to automate the detection of unusual objects and people. Anomaly detection can be made to work with varying degrees of automation. A system might be programmed, for example, to look for certain behaviors that are pre-defined as anomalous, such as running or moving erratically, loitering or moving against traffic, or dropping a bag or other items, and to perform detection of various kinds of violent behaviors such as fighting, punching, and stalking.[53] Other anomaly detection systems eschew that kind of supervised programming in favor of a "deviation approach" that let smart cameras learn on their own what is normal. Under that approach, AI agents are trained to crunch massive volumes of "normal" video of a particular scene and then consider new observations as abnormal or unusual if they deviate too much from the trained model.[54] In the streets of Johannesburg, the software iSentry, paired with facial recognition and webs of CCTV cameras, is programmed to detect and interpret "abnormal behaviour," pointing to the risk of automated forms of predictive policing.[55]

Analysis of crowds is a growing area of interest in AI as extensive research focuses on analyzing crowd behaviors and detecting anomalous or atypical behaviors.[56] Another direction in AI research consists in understanding "grouping in crowds" — figuring out who is socially connected with whom. One research team from Stanford University, for example, sets forth an approach for

---

[51] Ryoo M., 2011. "Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos". *IEEE International Conference on Computer Vision*. http://cvrc.ece.utexas.edu/mryoo/papers/iccv11_prediction_ryoo.pdf ; See also Baradel F., Wolf C., Mille J., and Taylor G., 2018. "Glimpse Clouds: Human Activity Recognition from Unstructured Feature Points". *CVPR.*
http://openaccess.thecvf.com/content_cvpr_2018/papers/Baradel_Glimpse_Clouds_Human_CVPR_2018_paper.pdf

[52] Ryoo M., 2011. see also Conner-Simons A., 2016. "Deep-learning vision system anticipates human interactions using videos of TV shows". *Phys.* https://m.phys.org/news/2016-06-deep-learning-vision-human-interactions-videos.html

[53] Ke S., Thuc H., Lee Y., Hwang J., Yoo J., and Choi K., 2013.; In fact, "Violence detection"— attempts to detect fights and similar behavior in videos — is an area of particular interest. One group, noting that it "may be extremely useful in video surveillance scenarios like in prisons, psychiatric or elderly centers," created a dataset of fight and non-fight videos to test violence-detection approaches. They concluded that "fights can be detected with near 90% accuracy." Bermejo E., Deniz O., Bueno G., and Sukthankar R., 2011. "Violence Detection in Video Using Computer Vision Techniques". *CAIP.*
https://www.cs.cmu.edu/~rahuls/pub/caip2011-rahuls.pdf;
Mohammadi S., Perina A., Kiani H., and Murino V., 2016. "Angry Crowds: Detecting Violent Events in Videos". *European Conference on Computer Vision.* https://link.springer.com/chapter/10.1007%2F978-3-319-46478-7_1

[54] Ke S., Thuc H., Lee Y., Hwang J., Yoo J., and Choi K., 2013.

[55] Kwet M., 2019. "Smart CCTV Networks Are Driving an AI-Powered Apartheid in South Africa". Vice News.
https://www.vice.com/en_us/article/pa7nek/smart-cctv-networks-are-driving-an-ai-powered-apartheid-in-south-africa

[56] Research on crowd analytics is also focused on behavioral prediction ("main directions, velocities, and unusual motions"), tracking a specific person through a crowd, and violence detection in crowds.
Junior J., Musse S., and Jung C., 2010. "Crowd Analysis Using Computer Vision Techniques". *Signal Processing Magazine, IEEE*. 27. 66 - 77. 10.1109/MSP.2010.937394.;
Hassner T., Itcher Y., and Kliper-Gross O. "Violent Flows: Real-Time Detection of Violent Crowd Behavior".
https://www.openu.ac.il/home/hassner/data/violentflows/violent_flows.pdf

"multiple people tracking" that takes into account the interaction between pedestrians using social as well as grouping behavior. The paper stresses the importance of using social interaction models for tracking in difficult conditions such as in crowded scenes.[57] Crowd analysis is an area that could be considered increasingly useful in epidemiological surveillances but it also raises the prospect of political and social surveillance of events like protests and rallies.

□ **Affect- and behavior-recognition**: The field of affect- and behavior-recognition is developing at rapid pace with algorithms being trained to identify what are purported to be the six basic emotions: anger, happiness, sadness, fear, disgust, and surprise.[58] "Pain intensity estimation" is also an active area of research, using a framework called the Facial Action Coding System (FACS), an attempt to encode all possible movements of facial muscles.[59]

There is already a significant market for emotion-recognition software — one forecast to reach at least $3.8 billion by 2025.[60] A company called Noldus, for example, claims that its deep learning- and FACS-based algorithms can identify those basic emotions, their intensity level, and a general measurement of happiness vs. sadness.[61] During the Sochi Olympics, Russian officials deployed video analytics software called VibraImage, which purported to detect agitated individuals by measuring facial muscle vibrations.[62] Oxygen Forensics, which sells data-extraction tools to clients including the FBI, Interpol, London Metropolitan Police and Hong-Kong Customs, announced in July 2019 that it also added emotion-recognition to its software, which includes analysis of videos and images captured by drones.[63] The company Affectiva says its product senses cognitive and emotional states, including levels of fatigue, distraction, anger, frustration, and confusion.[64] Affectiva is interested in quantifying emotions to get a complete understanding of individuals' overall wellbeing as a suicide prevention tool.[65] The company has also developed Peppy Pals,[66] a

---

[57] Leal-Taixe L., Pons-Moll G., and Rosenhahn B., 2011. "Everybody needs somebody: Modeling social and grouping behavior on a linear programming multiple people tracker". *IEEE International Conference on Computer Vision Workshops.* https://www.researchgate.net/publication/221430141_Everybody_needs_somebody_Modeling_social_and_grouping_behavior_on_a_linear_programming_multiple_people_tracker See also Alahi A., Ramanathan V., and Fei-Fei L., 2017. "Tracking millions of humans in crowded space in crowded spaces". *Group and Crowd Behavior for Computer Vision.* http://vision.stanford.edu/pdf/alahi2017gcbcv2.pdf

[58] Kosti R., Alvarez J., Recasens A., and Lapedriza A., 2017. "EMOTIC: Emotions in Context Dataset". *CVPR.* http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Lapedriza_EMOTIC_Emotions_in_CVPR_2017_paper.pdf.

[59] Martinez D., Rudovic O., and Picard R., 2017. "Personalized Automatic Estimation of Self-reported Pain Intensity from Facial Expressions". *CVPR.* http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Picard_Personalized_Automatic_Estimation_CVPR_2017_paper.pdf; Zhou G., Wu J., Zhang C., and Zhou Z., 2016. "Minimal Gated Unit for Recurrent Neural Networks". https://www.cvfoundation.org/openaccess/content_cvpr_2016_workshops/w28/papers/Zhou_Recurrent_Convolutional_Neural_CVPR_2016_paper.pdf

[60] Hegde Z., 2018. "Emotion recognition and sentiment analysis market to reach $3.8bn by 2025, says Tractica". *IoT-Now.* https://www.iot-now.com/2018/03/08/78263-emotion-recognition-sentiment-analysis-market-reach-3-8bn-2025-says-tractica/

[61] Leanne Loijens and Olga Krips, "FaceReader Methodology Note," 2018 white paper, Noldus Information Technology, available by request from Noldus, see https://www.noldus.com/human-behavior-research/products/facereader

[62] Herszenhorn D., 2014. "Heightened Security, Visible and Invisible, Blankets the Olympics". *The New York Times.* https://www.nytimes.com/2014/02/14/sports/olympics/heightened-security-visible-and-invisible-blankets-the-olympics.html

[63] Oxygen Forensics, 2019. "Detective 11.5". Oxygen Forensics. https://www.oxygen-forensic.com/uploads/press_kit/OF_RN_11_5_web.pdf

[64] http://go.affectiva.com/auto; https://youtu.be/V_rr7pDPdNM; a company called Nauto also offers products measuring driver distraction and other dangerous conditions to fleet operators: https://www.nauto.com/

[65] Affectiva's SDK project for suicide prevention, see: "SDK on the Spot: Suicide Prevention Project with Emotion Recognition." Affectiva, 14 August 2017. https://blog.affectiva.com/sdk-on-the-spot-suicide-prevention-project-with-emotion-recognition

[66] For more information on Peppy Pals, see: McManus A. 2017. "SDK on the Sport: Peppy Pals Education Apps Teaches Children SEL/EQ Skills." Affectiva, 23 March. https://blog.affectiva.com/sdk-on-the-spot-peppy-pals-educational-apps-teacheschildren-sel/eq-skills

series of education apps that teach children about social and emotional intelligence by learning from situations in an online world.

A growing amount of research focuses on detecting emotions via voice-analysis. Through voice-analysis, companies promise the ability to measures such things as energy, empathy, tone, and pace in a conversation.[67] Researchers are also studying "body affect analysis" — the ability to detect emotion by monitoring body movements. The underlying assumption is that the human body in motion provides a rich source of information about the intentions and goals of an actor, as well as about various aspects of his or her internal state.[68] Eye tracking is also a trending area of commercial and academic interest. Among the things that eye-tracking could be used to try to discover about a subject are intent, objects of interest, cognitive disorders, drug and alcohol use, and mental illness.[69] For instance, police in the US and the UK are using the eye-detection program, Converus, which examines eye movements and changes in pupil size to flag potential lies and deception.[70]

Combined with biosensors, algorithms are also able to recognize the physiological state of individuals, measuring signals such as heartbeat, vein flow, blood oxygenation, body heat and sweat. The combination of AI and sensing technologies allows to capture highly sensitive personal medical information, ranging from detection of arrhythmias and cardiovascular disease, to asthma and respiratory failures, physiological abnormalities, psychiatric conditions, or even a woman's stage in her ovulation cycle.[71] MIT Media lab, for instance, explores methods for nudging individuals towards well-being, parsing through their physiological, mobile phone and behavioural data.[72]

Several organisations have condemned not only the hype, but also the lack of rigorous scientific methods surrounding the current affect- and behaviour-recognition industry.[73] The problem is that, even without clear scientific efficiency and predictive value, affect-recognition tools may become harmful if applied to measure human emotional, cognitive and learning performance and experience in complex, critical domains such as health, education, employment, social welfare fraud-detection and policing.

As well explained by Harvard Professor Shosana Zuboff, we have entered a new tech era where human experience has become free material for behavioral analysis.[74] This emerging capacity is born out of the convergence between AI and technologies that capture the sensitive data of individuals' lives—their biometrics and bio-data, their movements and consumption patterns, their emotions and conversations. "The Internet of Bodies, Genomes and Minds," is how I call this new technological convergence that produces value out of the interference between human digital, physical and

---

[67] https://www.cogitocorp.com/product/

[68] Shields T., Amer M., Ehrlich M., and Tamraker A., 2017. "Action-Affect-Gender Classification using Multi-Task Representation Learning". *CVPR.* http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Tamrakar_Action-Affect-Gender_Classification_Using_CVPR_2017_paper.pdf

[69] Stanley J., 2013. "The Privacy-Invading Potential of Eye Tracking Technology". *ACLU.* https://www.aclu.org/blog/national-security/privacy-and-surveillance/privacy-invading-potential-eye-tracking-technology

[70] Harris M., 2019. "An Eye-Scanning Lie Detector Is Forging a Dystopian Future," Wired. https://www.wired.com/story/eye-scanning-lie-detector-polygraph-forging-a-dystopian-future/

[71] Rogers C., 2014. "A Slow March towards Thought Crime: How the Department of Homeland Security's Fast Program Violates the Fourth Amendment". *American University Law Review.* https://digitalcommons.wcl.american.edu/aulr/vol64/iss2/5/

[72] Umematsu T., Sano A., and Picard R., 2019. "Daytime Data and LSTM can Forecast Tomorrow's Stress, Health, and Happiness,". 41st International Engineering in Medicine and Biology Conference. https://www.media.mit.edu/publications/daytime-data-and-lstm-can-forecast-tomorrow-s-stress-health-and-happiness/

[73] Chen A. and Hao K., 2020. "Emotion AI researchers say overblown claims give their work a bad name". *MIT Technology Review.* https://www.technologyreview.com/2020/02/14/844765/ai-emotion-recognition-affective-computing-hirevue-regulation-ethics/

[74] Bridle J., 2019. "The Age of Surveillance Capitalism by Shoshana Zuboff review – we are the pawns". *The Guardian.* https://www.theguardian.com/books/2019/feb/02/age-of-surveillance-capitalism-shoshana-zuboff-review

biological data.[75] The terms "convergence" and "converging technologies" are chosen on purpose as it becomes difficult to understand powerful technological implications on populations by separating AI from big data, biometrics, the Internet of Things and its sensors, biology and genomics (which are a type of "population data"). As they are essentially digitized, converging technologies also operate in cyberspace, where other activities such as cybercrime or other technical domains such as cybersecurity and quantum computing will play a disruptive role, protecting or threatening the security of populations' data.

CONVERGENCE, DUAL-USE AND SOCIO-TECHNICAL SYSTEMS

- While designed for beneficial purposes, converging technologies exhibit functions that can easily be misused. This is why societal implications of these technologies need to anticipated and assessed before deployment in vulnerable context (underserved minority youth, deeply unequal societies). We need to develop a "theory of no-harm" and practice "socio-technical system" analysis.
- Such analysis cannot rely on self-regulation and corporate ethics alone. The global supply chains which produced converging technologies are complex and fragmented. We face a significant accountability gap and pervasive misalignment between ethics and business incentives.
- The private sector (data & tech) is invested in running and owning strategic elements of critical public infrastructure in health (hospitals, medical insurance companies) and education (schools, Ed Tech). There is a risk to lead to a form of "private automation and optimization" of human capital.

Dual-use technologies belong to a set of technologies that are conceived, designed and deployed for beneficial purposes but can also inherently cause harm, either accidental "unintended harms," or as a result of deliberate malicious intent. Converging technologies, which are today primarily developed by the private sector, could play a key role in increasing functionality across a wide spectrum of dual-use applications (including those falling in the hands of non-state violent actors through cybercrime). Yet, as they converge, technologies also become more automated and decentralised, blurring who is responsible for technologies' misuses, leading to what the author calls "atomised liability." Decision-makers in the public and private sector will therefore have to adapt, revise and upgrade governance models so that harms to populations can be mitigated even in an era of technological decentralization and automation, as well as in a context of distributed agency and liability. There is an urgent need to anticipate and devise how the private and public sectors will harness and regulate the dual-use potential of converging technologies. More sobering, it will become pressing to monitor how authoritarian regimes will be able to co-opt powerful private sector capabilities generated by technological convergence for surveillance, power and resource capture.

Converging technologies do not exist in a technical vacuum, but equally depend on the social, political and economic contexts in which they are developed, deployed and used. An array of factors and actors will influence if converging technologies meet the ethical expectations of vulnerable populations and whose interests and needs, they primarily serve. To better understand and track the complexities of converging technologies requires considering ways in which technologies are entangled in social relations, material dependencies, and political purposes.[76] Alongside such efforts, engineers and researchers from a range of disciplines need to conduct what is called social-systems analyses of technologies. They need to assess the impact of technologies on their social, cultural and political settings. A social-systems approach could investigate, for instance, how the use of a mobile phone

[75] Pauwels E., 2019. "The New Geopolitics of Converging Risks". United Nations University. https://collections.unu.edu/eserv/UNU:7308/PauwelsAIGeopolitics.pdf

[76] Jean-Christophe Plantin, Carl Lagoze, Paul N. Edwards and Christian Sandvig. "Infrastructure studies meet platform studies in the age of Google and Facebook," New Media & Society, 20, no. 1 (2018): 293-310, https://doi.org/10.1177/1461444816661553 ; https://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805

application which relies on AI and facial recognition to automate stunting diagnosis will impact local communities, in particular the relationship between field workers, children and parents. How will it impact the understanding of stunting at the household level? How will reliance on automation, optimization and AI emotional analysis subsequently influence strategies for accelerating human capital more broadly? Similar technologies are already in use to automate the diagnosis and support the management of mental disorders in underserved areas.[77] Again, interesting questions emerge about the impact on the doctor-patient relationship, trust dynamics within communities, personal data security and emotional manipulation.

The challenge at the intersection of converging technologies and human capital is to develop sufficiently powerful and systematic understandings of "technologies in societies/communities" to know where the potential lie for empowerment and responsible governance. Recently, the AI world witnessed a salient example of the gap between self-regulation, corporate ethics and meaningful accountability. Microsoft provided support and technology to an Israeli facial-recognition surveillance company called AnyVision that targets Palestinians in the West Bank.[78] AnyVision facilitates surveillance, allowing Israeli authorities to identify Palestinian individuals and track their movements in public space. Given the documented human-rights abuses happening on the West Bank,[79] together with the civil-liberties implications associated with facial recognition in policing contexts,[80] at a minimum, this use case directly contradicts Microsoft's declared principles of "lawful surveillance" and "non-discrimination," along with the company's promise not to "deploy facial recognition technology in scenarios that we believe will put freedoms at risk."[81] This is typically a problem of misalignment between AnyVision's actions and Microsoft's ethical principles in a decentralized AI supply chain.

Only by tracing across these sociotechnical layers can we understand the unintended ethical failures and potential misuses that could occur along the global supply chains of converging technologies. There are many sociotechnical data infrastructures needed for convergence and where unintended biases and discriminatory design could transpire: these include training data, test data, APIs, data centers, fiber networks, undersea cables, energy use, labor involved in algorithmic training and data-curation, and a constant reliance on specific technical expertise to develop and maintain the accuracy and safety of AI systems. Without an understanding of these complex technological supply chains, it is rather difficult, if not impossible, to advocate for and implement principles of fairness, accountability, and transparency.

---

[77] Maulik P., Kallakuri S., Devarapalli S., Vadlamani V., Jha V., and Patel A., 2017. "Increasing use of mental health services in remote areas using mobile technology: a pre–post evaluation of the SMART Mental Health project in rural India*". Journal of Global Health.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370210/

[78] Olivia Solon, "Microsoft Funded Firm Doing Secret Israeli Surveillance on West Bank," NBC News, October 28, 2019, https://www.nbcnews.com/news/all/why-did-microsoft-fund-israeli-firm-surveils-west-bank-palestinians-n1072116

[79] Human Rights Watch, "Israel and Palestine: Events of 2018," accessed November 21, 2019, https://www.hrw.org/world-report/2019/country-chapters/israel/palestine#1b36d4

[80] Evan Selinger and Woodrow Hartzog, "What Happens When Employers Can Read Your Facial Expressions?," New York Times, October 17, 2019, https://www.nytimes.com/2019/10/17/opinion/facial-recognition-ban.html

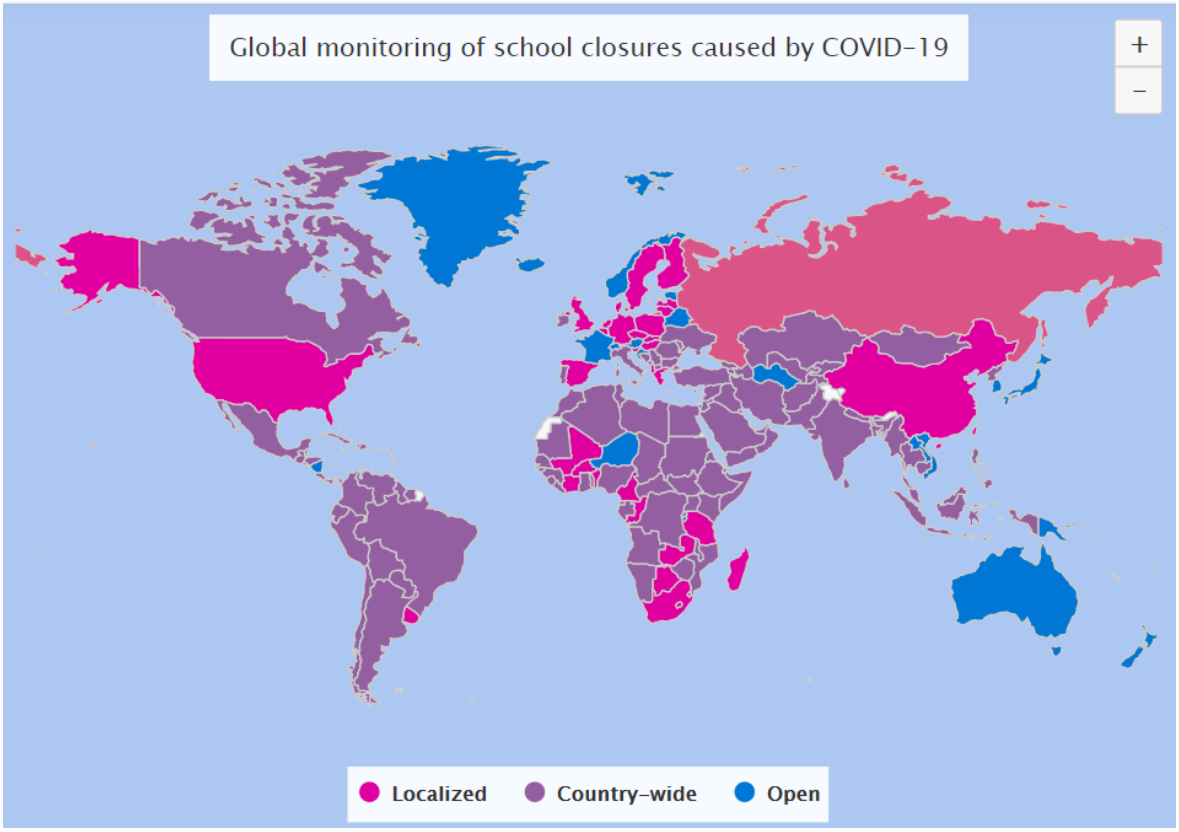
[81] Rich Sauer, "Six Principles to Guide Microsoft's Facial Recognition Work," Microsoft on the Issues, December 17, 2018, https://blogs.microsoft.com/on-the-issues/2018/12/17/six-principles-to-guide-microsofts-facial-recognition-work/

SECTION 3 – THREE IN-DEPTH CASE STUDIES FOR HUMAN CAPITAL

This section presents several case-studies [Ed Tech; Health/Nutrition; Democratized Innovation] that shed light on the dual-use potential of converging technologies for human capital, focusing on transformative opportunities but also specific risks to civilian and vulnerable populations.

CASE STUDY 1: AFFECTIVE COMPUTING FOR PERSONALISED LEARNING

The COVID-19 crisis has shut down traditional education programs throughout South Asia, from primary education to higher education. As reported by UNICEF, "many of the 430 million children affected by school closures in the region are now in danger of dropping out of the education system. Vulnerable and hard to reach children may never return to school if they get further behind due to not being reached with alternative ways to learn during school closures."[82]



Source UNESCO/Note: As of May 25, 2020, figures correspond to number of learners enrolled at pre-primary, primary, lower-secondary, and upper-secondary levels of education [ISCED levels 0 to 3], as well as at tertiary education levels [ISCED levels 5 to 8]. Enrolment figures based on latest UNESCO Institute for Statistics data. See methodological note.

In this pressing context, as education programs migrate "online," the World Bank and its partners will face increasing opportunities to "build back better" post covid-19 recovery. Converging technologies will present an array of opportunities to modernize education programs. In particular, AI and its potential to transform learning into a more adaptive and personalised process will feature high on the agenda. While the convergence of AI and data-capture technologies can be harnessed to accelerate human capital outcomes in the education sector, this prospect can also seriously undermine child rights and amplify children's insecurity in cyberspace.

[82] UNICEF, 2020. "Urgent need to secure learning for children across South Asia". https://www.unicef.org/press-releases/urgent-need-secure-learning-children-across-south-asia

AI EMOTIONAL LEARNING AS FRAMED BY TECHNOLOGISTS: Ed Tech companies are increasingly presenting AI as a technological method to optimize the real-time analysis of children's emotional and behavioural skills. The goal is to assist with personalised learning, development of social and emotion learning, to understand if students are struggling with class material, and which students need to be challenged further by class content. For instance, Affectiva is a leading facial emotion-recognition company that has shown a keen interest in education. Affectiva frames AI's innovative development for learning as follows: "What if we had intelligent learning systems that provided a personalized learning experience? Such a system would know you and your learning style, sensing your levels of engagement or frustration and then adapting in real time. It would offer a different explanation when you are frustrated, slow down a bit when you are confused, tell a joke when it's time to have some fun … just the way an awesome teacher would. These intelligent systems would be able to take action just at the right time with a personalized experience."[83]

In general, affective computing is a form of behavioural and emotional monitoring that relies on recognizing and sensing facial expressions, gaze direction, gestures and voice.[84] It can also encompass machines sensing and learning about heart rate, body temperature, respiration and the electrical properties of our skin, among other bodily behaviours. Being trained on classifications and labels, AI programs see with cameras, receive input features, register pixelated facial elements, process these against labelled training data and generate outputs of named emotional states. For example, computer vision may track the movement of lip corners, the speed with which this movement occurs, and the length of time corners of the mouth are moved from their usual position. While there exists a variety of emotional AI techniques, it is the analytics called "Facial Action Coding System" (FACS, cf. p 21) that is receiving the most interest from EdTech companies and legacy technology companies, such as Microsoft and Intel, that see opportunity in the education market.

With emotional AI applications, Ed Tech companies claim big promises, from identifying students in need of attention, helping diagnose the nature of insufficient sustained attention, predicting future lack of performance, to adapting learning materials. Yet, these applications, when scrutinized in their design functions, exhibit technical reductionism and weak scientific foundations, which is a problem for evidence-based education.

WEAK SCIENTIFIC FOUNDATION AND PREDICTIVE VALUE: First, they reduce complex emotional states to a basic coding of facial movements. For instance, Affectiva's product, Affdex, uses cameras in personal devices or public spaces to capture facial expressions of people as they view or interact with content or objects.[85] Affdex employs FACS-based analytics to assess the movement of 45 different facial muscles and key feature points, such as the eyes and mouth. The system then classifies arrangements of pixels from facial frames and translate them into named emotions. This result is essentially reduced to a taxonomy of facial movements. Yet, the power of Affectiva comes from its big data-capture, which, if unrivalled, sets the standard and contributes to the company's legitimacy: Affectiva has an ever more expansive facial biometrics collection with a database exceeding 8 million faces in 2019.

Second, a landmark 2019 review study found that efforts to "read out" people's internal states from an analysis of facial movements alone, without considering context, are at best incomplete and at worst entirely lack validity.[86] After reviewing over a thousand studies on emotion expression, the

[83] El Kaliouby R., 2017. "Announcing Affectiva Partnership with IBM Watson AI XPRIZE". *Affectiva*. https://blog.affectiva.com/announcing-affectiva-partnership-with-ai-xprize-why-emotion-needs-ai

[84] Lisa Feldman Barrett, Ralph Adochs, and Stacy Marsella, "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements," Psychological Science in the Public Interest 20, no. 1 (July 2019): 1–68, https://journals.sagepub.com/eprint/SAUES8UM69EN8TSMUGF9/full

[85] McDuff D., et al., 2017. "AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. *Affectiva.* http://www.affectiva.com/wp-content/uploads/2017/03/McDuff_2016_Affdex.pdf

[86] Idem Lisa Feldman Barrett, Ralph Adochs, and Stacy Marsella, "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements," Psychological Science in the Public Interest 20, no. 1 (July 2019): 1–68, https://journals.sagepub.com/eprint/SAUES8UM69EN8TSMUGF9/full

authors found that, although these technologies claim to detect emotional state, they actually achieve a much more modest outcome: detecting facial movements. As the study also shows, there is a substantial amount of variance in how people communicate their emotional state across cultures, situations, and even across people within a single situation. Of course, being first in collecting and mastering a critical mass of diverse facial data would confer companies in the space significant competitive value in sectors beyond education.

The lack of scientific validity and "portability" – how accurately can analytical tools be transferred from a domain to another – raises further questions about 1) how a taxonomy of facial movements becomes a valid representation of students' emotional experience in learning and interacting with peers; and 2) how it may lead children and older students to internalize AI-led codification of their own affective states. This could have a chilling effect on students as they could become self-conscious and paralysed at the idea of being emotionally "scored." While AI emotion taxonomies give a gaze of truth and quantification, they are primarily established and developed for purpose in adjacent sectors, including retail, advertising and criminal justice, where social scoring emerges as a new powerful instrument.

IMPLICATIONS FOR CHILD RIGHTS: The implications for child rights are substantial. Here, legal reflections should centre on issues of proportionality, data-purpose and data-minimization. In brief, is the pervasive collection of emotional, behavioural and soft biometric data about children necessary to achieve successful education? Given weak scientific foundations, is this form of data-collection even achieving the promised "personalisation" of education – when we know that one emotional taxonomy is used across populations and social, cultural contexts? The issue of human flourishing, social development and child educational benefits should also be recognized as they are in the UN Convention on the Rights of the Child.[87] The Convention clearly states the need to act in the child's best interests (Art. 3), the child's right to freedom of thought (Art. 14) and privacy (Art. 16), the right to develop full potential (§1 Art. 29), the child's right to liberty (§2 Art. 29), and the child's right to be protected from economic exploitation (Art. 32). Ed Tech companies might argue that AI will help children achieve their full potential, but such claim has to be supported by scientific valid arguments and has to provide guarantees that the above concerns about child rights are fully addressed.

WHAT NORMATIVE VISION? WHAT ARE TRAINING YOUTH FOR: Another argument coming from Ed Tech companies focuses on the fact that AI-led emotional applications constitute an invaluable tool to equip the next generation of students with social and emotional skills allegedly valued in the rising digital economy.[88] Such claim not only needs further assessment, but also a normative vision. First, are AI-led application performing better at developing creativity and resilience compared to peer-to-peer interaction and human mentorship? Or are these applications mainly providing social scoring that will be subsequently used in the labour market? Second, do we have enough clarity about the normative vision being proposed when it comes to integrating the next generation of students into future labour markets? Do we even know what we are "emotionally training" for? What we begin to envision is that there will be increasing competition for cognitive and creative performance between humans and machines. Therefore, is the commodification of behavioural and emotional data – early in the development of children – supporting machine or human performance, or both? Given AI systems' capacity for prediction, ubiquitous connectivity and updatability, we have to ensure that one type of intelligence is not drastically gaining competitive advantage at a pace and scale we cannot control.

PRIVATE OPTIMITZATION & AUTOMATION OF EDUCATION: AI trends in education also have to be understood in their broader political-economic context, the increasing privatisation, optimization and automation of public infrastructure, including schools. This trend is becoming apparent with the development of "smart and safe cities" where CCTV cameras and AI-led surveillance is delegated to the private sector.[89] Similar dynamic occurs when the World Food Program decides to rely on the US

[87] Office of the United Nations High Commissioner for Human Rights. 1990. Convention on the Rights of the Child. https://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf

[88] World Economic Forum, 2015. "New Vision for Education". https://widgets.weforum.org/nve-2015/

[89] AI Now Institute, 2019.

security company, Palantir, to manage biometrics ID systems and optimize food delivery for the most vulnerable populations across the globe.[90] Just like schools, hospitals and the broader health and medical industries are potential targets for privatization, optimization and automation.

In the field of education, a key recent development is the increased interest from legacy technology companies and large technological platforms. The education branch of the technology company Intel states that they are researching how recognition of affect may personalise learning experiences and provide adaptive learning.[91] Key factors for Intel are scope to recognise individual learners' expressions of excitement, frustration and boredom, so educators may modulate and optimise the difficultly levels of content. Intel also seek to track 'patterns of student motivation' and 'gauge emotional investment.' Microsoft Azure offers similar facial coding products which are marketed alongside their face identification and detection products.[92] Beyond sensing and detection, users of their Application Programming Interface (API) also grant capacity to search, identify and match faces within databases of up to 1 million people. These may be segmented by features, similarities and behaviour. Beyond class and school-level analytics, such face datasets can be used for behavioural analysis across much larger groups, potentially for wider social scoring purposes. Interestingly, Microsoft Azure is also the platform used for storing the facial and biometrics features of Indian children in the diagnosis of stunting (cf. Case-Study 2).

EDUCATION DATA & ALGORITHMS' OWNERSHIP: There are increasing signals of the privatization of education by large digital platforms. A major sign towards that direction is the fact that AI analytic systems operate on large, centralized, often private sector-owned cloud data centres (such as Microsoft Azure). Such trend raises critical questions of data and algorithms ownership, or "epistemic control." For instance, a company like Mindspark, which provides AI-based emotional analysis for education/tutorial programs, may assert ownership of the data the company collects about students' assessment.[93] If Google was to use its G suite for educational purpose, similar contentious issues about students' data ownership would likely surface. Recent studies have shown how digital platforms – such as Facebook and even Microsoft – have a poor record at protecting their datasets from being leaked or misused by third parties. Facebook, for instance, profits from sharing with other platforms, users' extremely personal information – such as heart beat and menstrual cycle data – obtained through smart applications.[94] An investigation by the Financial Times explains how Microsoft biggest facial data-sets was used by a range of US and Chinese corporations as well as by authorities in China to implement their automated ethnic profiling. "Both Sensetime and Megvii are Chinese suppliers of equipment to officials in Xinjiang, where minorities of mostly Uighurs and other Muslims are being tracked and held in internment camps."[95] Beyond issues of data ownership, the weakly regulated supply chains of AI create pervasive data-security threats and a growing accountability gap, with vulnerable populations – in this case, students – exposed, without consent, to potential privacy breach and data-based commercial exploitation. Under the same "privatization impulse," the Bridge International Academies (BIA) program has led to recent controversy over its plan to monetize the data it will collect about school children through their education career (cf. p 9-10).

---

[90] World Food Programme, 2019. "Palantir and WFP partner to help transform global humanitarian delivery". *World Food Programme.* https://www.wfp.org/news/palantir-and-wfp-partner-help-transform-global-humanitarian-delivery

[91] Intel Education. 2014. Emerging Technologies: Accelerants for Deep Learning. https://www.intel.com/content/dam/www/public/us/en/documents/training/emerging-technologies-lesson-plan.pdf

[92] Microsoft Azure. 2018. Face. https://azure.microsoft.com/en-gb/services/cognitive-services/face

[93] Kasinathan G., 2020. "Making AI Work in Indian Education". *Friedrich Ebert Stiftung.* http://library.fes.de/pdf-files/bueros/indien/15953.pdf p 6

[94] Dance G., LaForgia M., and Confessore N., 2018. "As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants". *The New York Times.* https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html

[95] Murgia M., 2019. "Microsoft quietly deletes largest public face recognition data set". *Financial Times.* https://www.ft.com/content/7d3e0d6a-87a0-11e9-a028-86cea8523dc2

REDIRECTING RESOURCES: Modern Ed AI/Tech remains a paradigm in which private interests shape public education. It also represents a new form of solutionism where AI and data-capture technologies – even when they lack scientific evidence – are supposed to solve structural problems in human capital (e.g. how teachers are trained, how schools are funded). One of the first consequences might be redirecting resources from training and upskilling teachers to investing in converging technologies and cutting-edge material equipment. Such prioritization may lead to epistemic asymmetry: who (which company, which demographic, in what cultural context) has the effective power to decide what is being taught and how? Who has a choice and who is being coerced in applying certain teaching methods and curricula? Who is being valued, trained and paid for mentoring the next generation of students? What are the implications for human capital at the community level? Increasingly, we may see epistemic methods used by teachers – where learning is a social activity, including peer interactions and mentorship – replaced by real-time and predictive AI analytics.

BIASES, POWER ASYMMETRY & INEQUALITY: Finally, promising personalisation at scale in packed classrooms is so tempting it might override the long list of concerns around children's data commodification. What about the right not to be emotionally profiled? What about the impact on relationships to peers, mentorship or emotional self-governance (chilling effect of being watched and scored in intimate dimensions)? As mentioned earlier in this paper, we have entered a new tech era where populations' data has become free material for behavioural analysis. Under an AI-led privatised Ed Tech paradigm, children' s behavioural and emotional experience risks being commodified not only to predict tomorrow's consumer markets, but to engineer behavioural demand for such future markets. In a context where the potential for socio-economic scoring is rising, the ability to control children's future behaviour has corrosive implications for agency and could lead to further inequality and disempowerment. For instance, Kasinathan explains how "in the Indian education scenario, the risk of bias is very high."[96] India already suffers from corrosive biases, ethnic and socio-economic stratifications which decides young individuals' futures. Such biases are so entrenched, they are inexorably mirror in any AI dataset collected about India's population. As eloquently said by Kasinathan, "Data sets embodying biases are bound to lead to predictive models that reinforce social discrimination, further impoverishing learning possibilities of students, especially those belonging to marginalised sections. This would bely the transformative potential of education in India."[97]

The above reflections have highlighted significant methodological, ethical, legal and normative concerns with emotional AI-led Ed Tech, especially applications based on facial coding. Below is an attempt at summarizing a few considerations that could be integrated into further technical and oversight assessment:

[96] Kasinathan G., 2020. p 8
[97] Idem

AFFECTIVE COMPUTING FOR ADAPTIVE LEARNING

□ **Scientific validity**: Does the technology achieve what its developers claim? Has the data been trained on a suitable diverse dataset? Has independent scientific scrutiny been solicited? What processes and checks and balances are in place to assess the predictive value and efficiency (on what criteria) when implementing emotional AI applications in education context?

□ **Universality**: Developers of facial coding rely on the assumption that basic emotions are universal or can be measured through a universal taxonomy, but what of 1) ethnocentric considerations; 2) representative training data; 3) social and cultural variation in emoting and learning; 3) individual-level variation in affective reactions, emoting and learning?

□ **Human-Machine Portability or what could be termed "technological reductionism:"** The existing emotional taxonomy is articulated in a way that makes sense to machines. Has consideration been given to the risk of creating an understanding of emotional life that suits data-analytics, but not people?

□ **Relationship with self, including self-reporting**: Has due diligence been assessed as to how these technologies may influence students' own emotional understanding? Will students be believed if they complain that they actually were paying attention or approaching learning from a different angle?

□ **Mental health and Discrimination**: Are ethical and policy safeguards in place if the technology detects signals of mental illness and how will these signals be balanced with privacy and data-protection concerns?

□ **Consent and Opting-out**: Power asymmetries and lack of access to or choice of schools render meaningful consent difficult or even impossible. What is the policy and rationale if parents do not want their children to be subject to predictive behavioural and emotional analysis by algorithmic programs? Has the trusting nature of people and their willingness to engage with non-human actors (and intelligence collection by non-human actors) been considered? Has the potential for social scoring been discussed in meaningful debates?

CASE STUDY 2: AI & CONVERGING TECHNOLOGIES FOR HEALTH DIAGNOSIS & SERVICE OPTIMIZATION

In the aftermath of the covid-19 crisis, international and humanitarian organisations will be tempted to shift to health services and diagnostics processes that are automated, remote ("no-touch application") and optimized by algorithms. This shift has for implication to transfer the quantitative but also the qualitative analysis of insights related to health issues to machine intelligence instead of human intelligence (acquired and performed by local doctors and field workers). The phenomenon of "automation bias" – the natural tendency for humans to favor suggestions from automated decision-making systems and to ignore contradictory information – significantly raises the stakes for the use of AI in sensitive health diagnoses. In addition, due to the volume of raw data, it may be impossible to quantify false negatives – i.e. what the system is missing. Such shift will have broader implications for how well human workforce is trained and prepared to understand and identify health issues at the household and local community levels.

STUNTING - Accurate Intelligence for remote diagnosis of malnutrition?

Stunting remains a major global health challenge, with irreversible impairment in a child's physical, education and cognitive development occurring in millions of children due to lack of effective interventions. Impaired linear growth is also linked with cognitive impairment, resulting in lower attained schooling and educational performance, leading to reduced productivity and decreased income-earning capacity,[98] and significantly greater risk of developing chronic non-communicable diseases such as cardiovascular disease and diabetes in adult life.[99]

Complex, Multi-factor Diagnosis: Stunting is associated with a multitude of determinants, but the underlying biological pathways remain unclear.[100] Potential risk factors include maternal ill health, intrauterine growth retardation and low birth weight, infant undernutrition, recurrent exposure to infectious diseases and enteric pathogens, micronutrient deficiencies, as well as socioeconomic influences (such as maternal education, paternal employment and exposure to environmental pollutants).[101] The relative importance of these risks varies between settings depending on socioeconomic status, diet and infectious disease profile[102] and attempts to identify common predictive factors have thus far had limited success.[103]

The process of diagnosing stunting usually relies on standardized measurements and physical exams but recently these practices have been suspended in many developing countries in the aftermath of the covid-19 crisis. The below application based on AI and sensing technologies provides a remote, no-touch solution.

Potential Solution Relying on Technological Convergence: AI (machine learning), facial- and body-recognition for data-analysis of facial and body scans; A mobile application that incorporates infrared measurements, a camera to capture scans and videos, and augmented reality for producing 3D scans; Cloud Data Processing and Storage

---

[98] Black R., et al., 2008. "Maternal and child undernutrition: global and regional exposures and health consequences". *The Lancet.* https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(07)61690-0/fulltext

[99] Barker D., 2007. "The developmental origins of chronic adult disease. *Acta Paediatrica.* https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1651-2227.2004.tb00236.x

[100] Prendergast A. and Humphrey J., 2014. "The stunting syndrome in developing countries". *Paediatrics and International Child Health.* https://www.tandfonline.com/doi/full/10.1179/2046905514Y.0000000158

[101] Fawzi M. et al., 2018. "Lifetime economic impact of the burden of childhood stunting attributable to maternal psychosocial risk factors in 137 low/middle-income countries". *BMJ Global Health.* https://gh.bmj.com/content/4/1/e001144

[102] de Onis M. and Branca F., 2016. "Childhood stunting: a global perspective". *Maternal and Child Nutrition.* https://onlinelibrary.wiley.com/doi/full/10.1111/mcn.12231

[103] Prendergast A. and Humphrey J., 2014.

□ Microsoft and Welthungerhilfe, one of the largest private aid organizations in Germany, are testing "Child Growth Monitor," an AI-powered mobile phone application that can capture and analyse children's facial and body scans, sometimes videos, to diagnose cases of stunting and malnutrition in three states in India (Madhya Pradesh, Maharashtra and Rajasthan).[104]

□ A U.S. data-analytics company, Kimetrica, is working on similar method for remote, automated, predictive diagnosis of stunting in areas affected by conflict or humanitarian emergencies.[105]

How the Current Diagnosis Problem is Framed by Technological Actors: The process of diagnosing stunting and assessing the nutritional status of children usually involve field workers weighing the children and measuring different body parts like circumference of the head and upper arms. Such an effort requires a significant workforce of field practitioners to be deployed across large areas to conduct health visit and physical checks of children aged between 6 months and 5 years. The technological developers of "Child Growth Monitor" associate current failed attempts at diagnosing stunting with a problem of lacking accurate data, performing improper data curation leading to weak analytical and predictive results. Most field workers seem to be ill-equipped or unskilled to properly collect, manage and analyse data as they write measurements manually on paper, save these measurements in log books and then transferred them to Excel sheets. This is alleged to be a lengthy, time consuming process, prone to human errors, that result in lack of predictive insights to help at-risk children. In addition, the high potential for errors in data collection prevents development of accurate, real-time visual geospatial maps where acute cases of stunting could be located to inform aid workers and humanitarian agencies and optimize health delivery services (advanced medical care and therapeutic food).

Functional Augmentation Provided by Converging Technologies: The smartphone mobile application, "Child Growth Monitor," uses facial and body-recognition technology equipped with a set of infrared sensors that can capture 3D measurements of a child's height, body volume and weight ratio, as well as head and upper arm circumferences down to the millimetre. In addition to facial and body scans, personal data collected for each child includes family name, name of guardian, name of child to scan, age, sex, birthdate and location. A guardian has to sign a consent form, be present and sometimes capable of using the smartphone application to take front, back and 360° snapshots and videos of the child to be diagnosed. When mobile connection allows the scanning process to be online, the data-capture takes a few minutes and measurement data become available in real-time on Microsoft Azure cloud platform. Remote nutritionists and IT specialists evaluate the scans by using the algorithmic platform "Microsoft AI solutions," to assess the health and nutritional status of children. Then, the diagnosis appears on the mobile application. If there is not enough mobile connectivity in the household/location, the scanning process happens offline and the children's facial and body scans have to be uploaded on the cloud platform later on. In case of low or no mobile connectivity, the diagnostics cannot be delivered and explained to the guardian or parent in real-time. Each diagnosis is used as training data to ensure that the algorithmic system gets more precise and more intelligent with each set of measurements.

To achieve high accuracy in diagnosis, the algorithmic training set needs to be large, diverse, and accurately annotated. In other words, the algorithmic system requires extensive capture of facial and body scans from subjects from different ethnic origins. To train the algorithm, the developers are therefore considering an alternative technique to labelling huge amounts of existing data by producing synthetic images ("rendered data") from a simulator.[106] Yet, training machine learning models on standard synthetic images is problematic as the images may not be realistic enough, leading the model to learn details present only in synthetic images and failing to generalize well on real images. One

[104] https://github.com/Welthungerhilfe/ChildGrowthMonitor
[105] https://kimetrica.com/services/modeling-and-simulation/
[106] Allen B., Curless B., and Popovic Z. "The space of human body shapes: reconstruction and parameterization from range scans". http://grail.cs.washington.edu/projects/digital-human/pub/allen03space-submit.pdf

approach to bridge this gap between synthetic and real images would be to improve the simulator which is often expensive and difficult, and even the best rendering algorithm may still fail to model all the details present in the real images. This lack of realism may cause models to overfit to 'unrealistic' details in the synthetic images. To help close this performance gap, the developers are exploring methods for refining synthetic images to make them look more realistic.[107]

Photo 1: "Child Growth Monitor" in Action (Source)

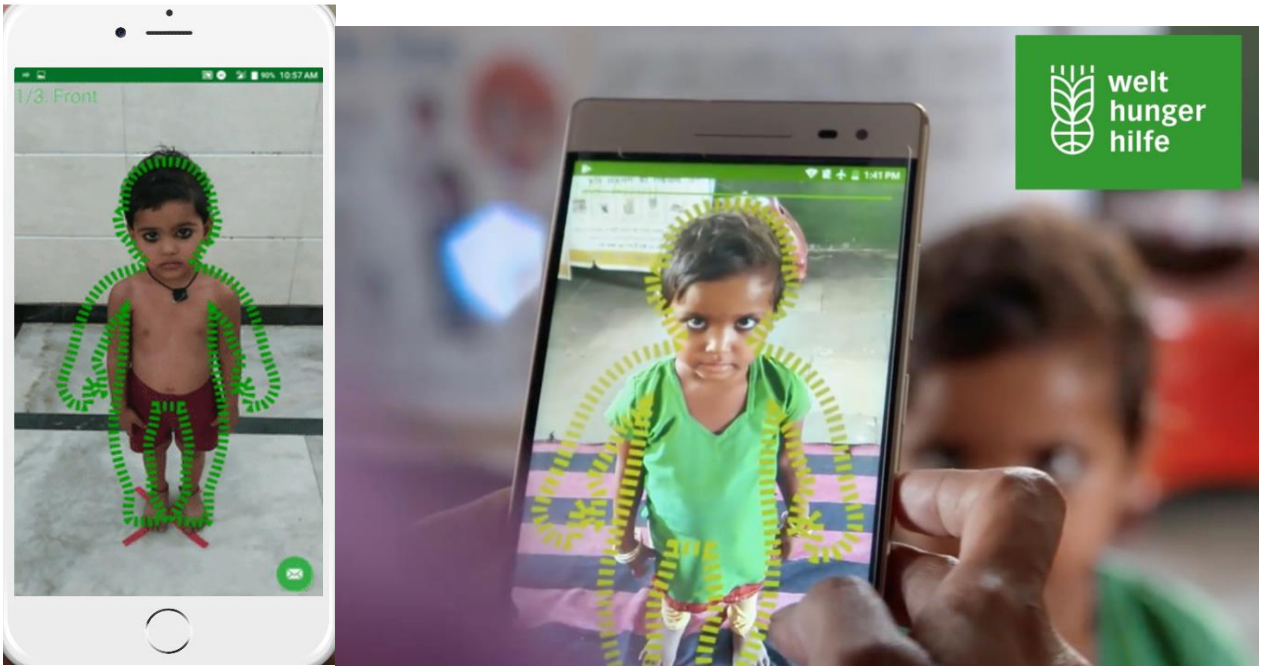


Photo2 (Below): Synthetic Data (Source)

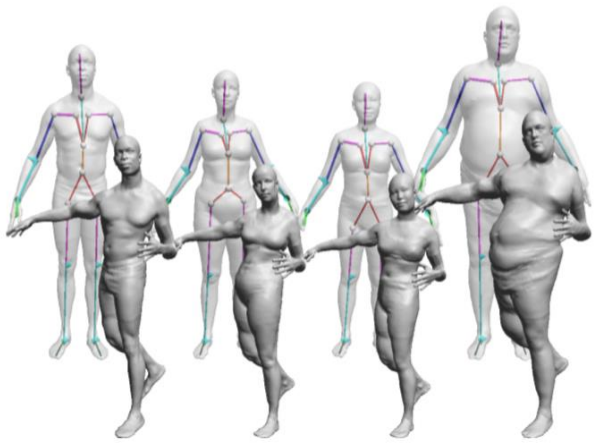


Figure 9: Skeleton transfer. We manually created a skeleton and skinning method for the scanned individual in the top left. The skeletons for the other three scanned individuals in the top row were generated automatically. In the bottom row, we show each of the parameterized scans put into a new pose using the skeleton and transferred skinning weights.

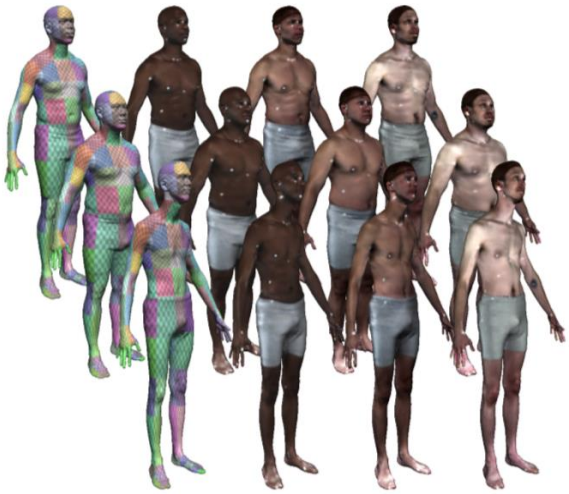


Figure 6: To test the quality of our matching algorithm, we apply the same texture (each column) to three different meshes. The mesh in each row is identical. On the left, we use a checkerboard pattern to verify that features match up. The right-hand 3×3 matrix of renderings use the textures extracted from the range scans. (The people along the diagonal have their original textures.)

[107] Shrivastava A., Pfister T., Tuzel O., Susskind J., Wang W., and Webb R., 2017. "Learning from Simulated and Unsupervised Images through Adversarial Training". *arXiv.* https://arxiv.org/abs/1612.07828

Photo3 (Below): Synthetic Data (Source)

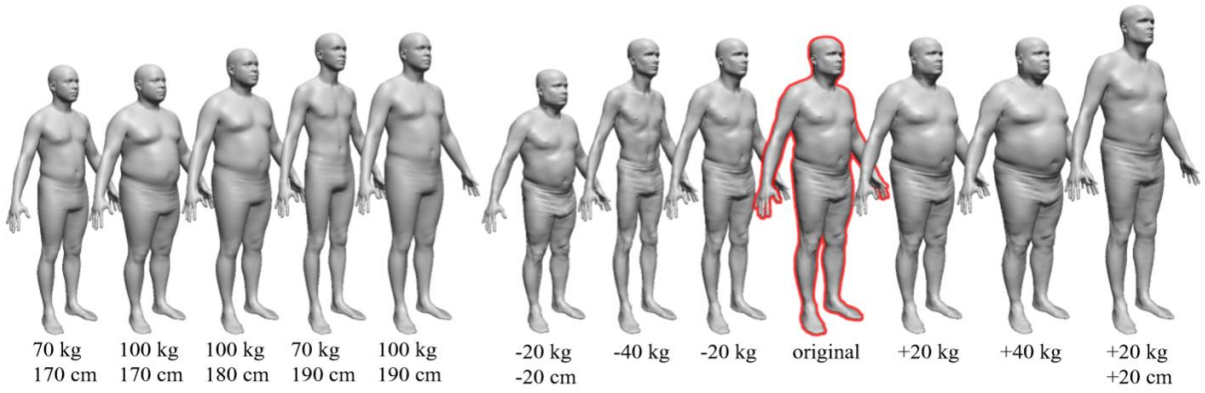


Figure 10: The left part of this figure demonstrates *feature-based synthesis*, where an individual is created with the required height and weight. On the right, we demonstrate *feature-based editing*. The outlined figure is one of the original subjects, after being parameterized into our system. The gray figures demonstrate a change in height and/or weight. Notice the double-chin in the heaviest example, and the boniness of the thinnest example.

Similar work is also undertaken by a US data-analytics company, Kimetrica, which is using facial-recognition and algorithmic analytics to diagnose stunting in children in Kenya. The mobile application and underlying algorithmic process is called "Methods for Extremely Rapid Observation of Nutritional Status" (MERON). Kimetrica achieved proof of concept with MERON for children and a preliminary classification accuracy level of 60 percent, using 3,500 images of children under-five (6-59 months). The next step for product development is a significant increase in its accuracy for malnutrition detection in children under-five from 60 percent to over 90 percent, which will require collecting additional image data.

● **Expected Benefits**: Using converging technologies to automate the remote diagnosis of stunting in young children is presented by the project developers as having multifaceted benefits.

□ **Automation**: An increase in the accuracy of collecting data on malnutrition; Use of inconspicuous measurement tools and a less invasive method to measure malnutrition; Easier data collection in hard to access, high risk or conflict areas, and areas where physical handling of children is culturally not acceptable; A no-touch, remote solution during the covid-19 crisis

□ **Redirecting resources** (resources used for training enumerators to take accurate weight for height measurements): Significant amounts of funds are currently used to support field workers who conduct in-person physical diagnosis and manual measurements of stunting. This type of funding could be redirected for other humanitarian purposes, such as future reinvestment in children's quality of life.

● **Potential Risks**: The implications of transitioning from physical exams done by field workers and doctors to an automated, remote system for diagnosing stunting has significant implications for households, communities and health workers.

□ **Accuracy of Datasets, Scanning Process and Predictive Value**: For all technologies that use AI embedded within diagnostics, it is important to ensure that: 1) the training dataset is large enough, is of quality and representative to the target population; 2) the predictive analyses produced by algorithms are accurate, precise and reproducible. What training data was used for the algorithm? Has algorithmic bias been systemically assessed? Has the reproducibility of the model been measured? How often will the algorithm and the training data be updated? Have standards been used in documenting the model learning process to avoid the black box effect[108] (a system whose inner workings are not visible)?

[108] "Black box" is a metaphor describing a system in which the input data and the results are known but the process that leads from one to the other is not visible.

While the "Child Growth Monitor" application is already being used in the field, most of the above questions still need to be fully addressed by the tech developers. The current dataset for the diagnostic system includes gold standard anthropometric measurements of height, weight, MUAC and head-circumference from 12 healthy children, which amounts to more than 700 3d scans of those children (~240 complete front/360°turn/back scans). In the next sixth months, the developers aim to collect data from about 10.000 children of which around 40% will be MAM and 2-4% will be SAM. Assessment about potential biases, reproducibility and standardization is not yet public or available to end-users as it seems to be a work in progress.[109]

At the intersection of AI and diagnostic research, issues usually abound about the accuracy of algorithmic design and the hidden biases that could corrupt health decision-making if training data sets do not carefully mirror the health determinants of local populations. For an algorithmic system to make a fairly complex diagnostic such as stunting, the system will need to be trained with very large datasets including subjects of diverse ethnic backgrounds that represent the population's heterogeneity. Such challenge of not only acquiring, but also labelling massive datasets, explains why, in this case-study, the tech developers are exploring the potential to rely on synthetic data - rendering 3D-parametrized morphological models (Photo 2 & 3) – to train their algorithm about what "normal" and "abnormal" body shapes look like in children aged from 6 months to 5 years. Training algorithmic systems with synthetic data raises the question of whether these artificial datasets can reflect the full range and depth of stunting symptoms, including less obvious health signals that appear in living bodies, not digital bodies.

Noise and morphological biases may also intervene during the scanning process where lightening might not be optimal in poor living conditions. Currently, the accuracy of one of the applications - the MERON predictive diagnostic – is only about 60%, which is still quite low. Accuracy of predictive value is not provided for the other application. In the "Child Growth Monitor" case, the automated analytical process also leads to a fairly reductionist set of result: "normal," "MAM," and "SAM," the two last results being associated with a recommendation to seek care at Anganwadi centers.

□ **Data Security, Privacy and Consent**: The assemblage of AI, facial- and body-recognition, and infrared sensors will capture sensitive morphological measurements, health and personal data. As it is also equipped with video capture, the diagnostic tool could include in the near-future cognitive and other behavioral data. Yet, one reason evoked by developers for adopting automated diagnostic is to avoid using conspicuous, invasive methods in underserved and conflict-prone areas. In this context, ensuring data-security, privacy and informed consent is paramount. Does the technology have a privacy policy for any data gathered? Are the data generated by the technology encrypted? Who will have access to the data and who will own the data? Where and how will data be stored? How will data security be managed including in the event of a breach? How will data be recovered in the event of loss?

As some households may not have connectivity, the facial and body scanning process has to be working online and offline. Data transfer and storage therefore involves image scans (not just anonymized data-points) and personal information. Personal data are encrypted and stored on the mobile application's backend. The morphological scans are uploaded to Microsoft Azure/cloud. The time for data retention is supposed to be limited, but is not given. When it comes to data-access by third party, access is provided on as-needed basis for other humanitarian actors interested in building similar applications.

While the tech developers have conceived basic data protection mechanisms, image scans, videos and personal data of children still depend on the security level offered by the mobile application's backend, digital techniques used for data-transfer and the cloud computing platform. Microsoft Azure and AI cloud solutions are deemed secure by the tech developers but such assessment might not be sufficient as cloud data-centers are increasingly targets for cyberattacks. Personal data, facial and body scans are at risks for data-manipulation and data-exfiltration to third-party.

[109] https://github.com/Welthungerhilfe/cgm-ml

Data-manipulation – also called "data-poisoning – is becoming part of the threat landscape for health and critical information infrastructure. For instance, researchers in Israel successfully trained algorithms to hack hospital CT scans.[110] By generating false lung tumours that conform to a patient's unique anatomy, this algorithmic experiment led to a misdiagnosis rate in excess of 90%. Researchers at Harvard University have tested adversarial attacks against algorithms used to diagnose skin cancer images, demonstrating that such attacks require only modifying a few pixels in the original biopsy picture to corrupt the diagnosis.[111] These are two examples of how autonomous malware can automate the manipulation of medical datasets, expanding the cyberattack' surface to the health and diagnostics industries.

If data-theft happens, it could lead to different forms of unintended misuse, from stigmatization, targeting of communities in conflict-prone areas to online data-abuse (harassment, pornography). Recent investigations have also pointed to the fact that large tech companies, start-ups and universities have leaked face datasets to other parties before. In 2019, the Financial Times reported that Microsoft had quietly pulled from the internet a database of 10 million faces, which had been used to train facial recognition systems around the world, including by military intelligence groups and Chinese firms such as Alibaba, SenseTime and Megvii. The people whose photos were used were not asked for their consent as the images were simply scrapped from internet platforms. This is not a problem isolated to the private sector. The FT investigation also reports that two other datasets from academic institutions – Duke and Stanford universities – had also been built using data scrapping and without informed consent.

In light of rising data-security threats and given the application's information life-cycle, there might be a need to anticipate and improve the following inflection points: could data-sets be accessed in the future by state authorities, insurance industry, commercial interests? What is and will be the exact time limit for children's data-retention? How to ensure informed, voluntary consent by vulnerable, under-skilled guardians/parents when relying on complex technologies and data architecture? These issues will need to be better defined and strictly implemented.

□ **Technological Ownership, Transfer and Benefit-Sharing**: Structural inequalities and vulnerabilities could be on the rise if AI health services are mainly provided by private companies outside of the local innovation ecosystem. A fundamental question centers around the required mechanisms that can allow for transfer of technology, skills and data to local doctors and hospitals. Who will implement, use and maintain the automated diagnostic system? Is this system compatible or interoperable with other diagnostic and health information systems? A question pertains to who has access to smartphones that can be equipped with 3D and infrared measurements application. Second, how is the required training for proper use provided and to whom?

The current business model of the digital economy incentivizes corporations to monetize large amounts of behavioral data about populations. For private sector companies, datasets of subjects matching diverse ethnic subgroups are a trove that allows to train software with more precision and acquire competitive advantage in an array of future markets. Yet, what are the benefit-sharing mechanisms for populations giving away their children's faces, helping train algorithmic systems of big companies? How are these personal datasets helping build future markets and groups of customers?

□ **Reductionism & Solutionism-The "Remote Effect:"** Stunting does not end with its remote diagnosis. What are the broader ethical and societal implications at the household and community levels, to the use of an automated, "no-touch" algorithmic system for diagnosing stunting?

[110] Mirsky Y., Mahler T., Shelef I., and Elovici Y., 2019. "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning". *arXiv.* https://arxiv.org/abs/1901.03597

[111] Finlayson S., Bowers J., Ito J., Zittrain J., Beam A., and Kohane I., 2019. "Adversarial attacks on medical machine learning". *Science.*
https://science.sciencemag.org/content/363/6433/1287.full?ijkey=OXnSsEp.Iagl6&keytype=ref&siteid=sci

Stunting involves a complex interaction of genetic, household, environmental, socio-economic and cultural influences. Automating diagnosis for stunting based primarily on remote morphological measurements (without physical exam) may lead to missed opportunities for identifying and understanding other medical factors at play in children' s poor development. Other important health, but also cognitive, emotional and relational factors may be missed as well as other household factors (related to environment/toxicity, sanitation, gender violence, etc.) that could be detected by field workers.

Automated diagnosis might become the norm. In the context of the covid-19 crisis, manual measurements and physical exams of children have been suspended because it became impossible to keep a safe distance. Automation may lead to reducing the size of and the support to the field workforce that usually conducts health visits with unintended, harmful consequences. One of these consequences could be a lack of training of field workers about what healthy morphological and cognitive development means, how it materializes in a healthy child, and what factors account for creating positive child's growth conditions. Less trained field workers may prevent the needed, expected behavioral change by members of the household when it comes to nutrition, health and cognitive development. What are the larger unintended – or un-anticipated – implications at the household and community levels?

CASE STUDY 3: LIFELONG LEARNING & MENTORSHIP IN DEMOCRATIZED INNOVATION ECOSYSTEMS

Source: Mechamind, Dhaka.

As the World Bank builds its own innovation capacities – for instance, the ITS lab which focuses on foresight, design and system thinking – the institution could also play a role in supporting and collaborating with the global networks of democratized innovation ecosystems that currently thrive and flourish in San Francisco, Shenzhen, Yucatan, Kumasi to Mumbai and Kathmandu.

In Kumasi's Lab, young innovators build drones that can deliver vaccines to isolated health centres.[112] Similar democratized innovation is happening in Kathmandu, Nepal where local engineers got trained by the start-up Fusemachines to build drones from scratch and deploy a low-cost delivery system for health equipment.[113] In Mexico's Lab, *Interspecifics*, inventors have designed algorithms that can recognize signals and interactions between micro-organisms inside biological cultures – a type of research that could serve better microbiome-based health diagnosis and ecosystem management.[114] In Shenzhen's Open Innovation Lab[115] mashups of self-made engineers, inventors and artists are working on projects ranging from drone swarming for diagnosing diseases on crops to precision algorithms for companion robots.

"AI FROM THE STREETS" – DEMOCRATIZED LEARNING IS A SOCIAL EXPERIENCE: The real strength of democratized innovation ecosystems, such as Maker Spaces and Community Labs, goes beyond the technical; it is social. What drives innovation in these spaces involves a common ethos characterized by de facto interdisciplinarity, peer-to-peer knowledge-sharing, increased self-esteem, acceptance and empathy. This is why they provide an innovative and alternative incubator for individual and collective empowerment using cross-discipline synergies. Moreover, this form of empowerment is not only open to students but also to STEM field teachers, providing them with a sustainable exposure to emerging technologies that can contribute to more formal classroom teaching. Community labs offer "libraries" of technologies and equipment to be borrowed or lent.

TRAINING GIRLS, MINORITY YOUTH & TEACHERS: Community Labs offer the general public access to tools (3D-printers, laser cutters, DNA sequencers, etc.), technological know-how and mentorship in diverse areas of expertise from engineering, programming, robotics, digital fabrication and biotechnologies to basic digital literacy. These community spaces present huge potential to provide effective STEM education, exposure, and mentoring activities to underserved minority youth. In Baltimore (USA), the Digital Harbor Foundation (DHF) provides an exceptional example of a community lab that, beyond structured educational programming, supports youth with access to meals and transportation.[116] DHF does not only integrate more than a thousand underserved minority students a year, but it also provides professional development opportunities to build teachers/educators' capacity to teach technology skills and integrate tech learning, innovation, and creativity into their own learning environments. In 2015, DHF served 1,381 K-12 youth and 300 educators with direct-service technology and engineering programming ranging from 3Dprinting to human-centred design. 76% of DHF youth served live under Federal Poverty Line standards, while 73% are minorities and 50% are females. How can such a model be scaled up to other cities and regions? What if cities could be globally connected, yet locally inventive and inspired by a diversity of knowledge and vision?

INCENTIVIZING MENTORSHIP FROM START-UPS, PROVIDING CERTIFICATES FOR AI SKILLS: Educating engineers from across the globe in converging technologies will help increase the diversity of tech talent. Someone who experiences complex problems in his/her own country could be more suited to try and solve those problems with AI. For example, a Nepali engineer who wants to use machine learning to predict crop yields of their community will be better informed about Nepal's farmlands than a graduate from Silicon Valley. Young students and engineers in remote developing countries also have the ability to perform -- and, at times, outperform -- the ones who have degrees from elite institutions in the West. There is untapped talent in places often neglected to local detriment.

But how do we train young local engineers in far-flung places to build drones, robots and complex systems? One answer may be a combination of online resources/courses and significant mentorship and peer-to-peer on-site training. For instance, the two MIT courses – How to Make (Almost)

---

[112] See the Kumasi Hive website, available at: https://www.globalinnovationexchange.org/organizations/kumasi-hive

[113] Maskey S., 2018. "The dream of democratizing AI takes flight". *Miratech.* https://miratechgroup.com/about/news/the-dream-of-democratizing-ai-takes-flight-local-talent-developing-medical-drone-delivery-in-nepal/

[114] See Interspecifics' website, available at: http://www.interspecifics.cc/-/about/

[115] See Shenzhen Open Innovation Lab's website, available at: https://www.szoil.org/

[116] https://www.digitalharbor.org/whatwedo/youth/

Anything[117] (focused on fabrication with digital and physical tech) and How to Grow (Almost) Anything[118] (focused on synthetic biology with digital and biotech) – are given online through virtual workshops where concept and techniques of fabrication are explained. Like a network with several regional and local nodes, students and engineers across the world first take part in the classes, and then learn to teach some of these classes as their skill set starts improving.

An interesting example is the start-up Fusemachines that launched a fellowship program that allows students in Nepal to develop high-level skills in programming and solving machine learning algorithms -- eventually leading to a Masters in Artificial Intelligence from Columbia University.[119] The mix of an online course available to the community-lab which provides on-site mentorship and peer-to-peer guidance works very well with students. With this model of learning, several engineers from Nepal have graduated with certificates in AI from Columbia University and have been paired with AI companies, thus helping to reduce the opportunity gap in tech jobs. The AI learning program has still to be scaled up but has already expanded to three additional locations: the Dominican Republic, New York City and Rwanda.

A third example is the one provided by a young woman I met in a community lab in San Francisco that benefited from mentorship and peer-to-peer training.[120] Growing up, Elodie witnessed her brother suffering from sudden crises called pneumothoraxes, triggered by a disease in which a lung collapses and separates from the chest wall. In severe cases, doctors resort to creating scar tissue on the wall as a grip to stitch back the lungs, an invasive treatment. But for Elodie, it was too painful and too slow. So, with the help of a mentor and other peers interested in biotech, she came up with her own, less invasive design—a "biological Velcro" that would leverage the inner mechanisms of proteins to bind her brother's outer lungs to his pleural cavity. For weeks, Elodie dissected the literature to find the proteins that are responsible for helping cells bind together. After narrowing down her search to a few candidates, she genetically modified them to enhance the binding effect. She made sure her proof of concept was reproducible, obtaining three optimally engineered proteins that bind very tightly to lung cells. Soon she will start bio-printing the engineered proteins on a "molecular patch," a thin matrix of collagen to be placed between the chest and the lungs. In collaboration with a university research team – she benefited from her community lab's and mentor's connection – she will then explore opportunities for clinical testing.

This is not science fiction. Biotechnologies have progressed to a point where it is now possible for young engineers and students to be taught how to use gene-editing techniques, which aim to modify the genetic code underlying cells and proteins. Advances could be unprecedented with the next generation learning how to turn their own ideas and know-how into new bio-constructs. Just like algorithms in software engineering, our cells have become intelligent-design material.

SERVING LOCAL NEEDS, EMPOWERING LOCAL YOUTH: Democratized innovation ecosystems are ready to learn, adapt and harness AI and converging technologies for their own local needs. Now, imagine scaling up democratized innovation ecosystems where AI engineers are incentivized to perform in-person (and online) mentorship for the next generation of young algorithms' designers in Dhaka and Islamabad. Imagine a democratized innovation lab where a young woman can be certified for new skills that add value to computation, where underserved users can experiment with new technological designs. Such a vision could be scaled up with the support of the World Bank and in

[117] Gershenfeld N., 2012. "How to Make Almost Anything". *Foreign Affairs.* https://www.foreignaffairs.com/articles/2012-09-27/how-make-almost-anything; https://ocw.mit.edu/courses/media-arts-and-sciences/mas-863-how-to-make-almost-anything-fall-2002/

[118] https://bio.academany.org/

[119] Crane L., 2017. "NYC company, with an assist from Columbia online teaching, is training AI talent in places few other companies think to look". *Columbia University.* https://www.cs.columbia.edu/2017/a-nyc-company-with-an-assist-from-columbia-online-teaching-is-finding-ai-talent-in-places-few-other-companies-think-to-look/

[120] Pauwels E., 2018. "The rise of citizen bioscience". *Scientific American.* https://blogs.scientificamerican.com/observations/the-rise-of-citizen-bioscience/

collaboration with schools, universities and the mentorship of tech engineers and start-ups. A model like the one proposed by Fusemachines could be studied and adapted to other national contexts and underserved needs.

The global innovation ecosystem could be slowly transformed if community labs continue to export their democratized approach to AI and converging technologies. The future of AI convergence could be defined, invented and implemented by bottom-up networks of inventors and engineers, rather than large corporate platforms alone. In this context, the World Bank and international development partners should discuss how to empower and oversee democratized innovation ecosystems, the "grassroots," in their effort to design and deploy AI and converging technologies for solving local social problems.

To foster inclusiveness, the World Bank could help support mentorship and certification programs, as well as fellowship exchanges between democratized ecosystems, private companies, start-ups and the WB/UN Innovation Labs. The ultimate goal would be to foster inclusive forums of engagement in which to share lessons learned, exchanges of innovation and oversight practices for AI convergence.

Source: Mechamind

One region in India that has built on the powerful global networks of fabrication labs is Kerala. Not only, Kerala has nurtured a set of fabrication labs where students can develop problem-solving skills, the region also acquired a super fab-lab that will collaborate and connect the whole local lab network with MIT Centre for Bits and Atoms.[121] Kerala's network of fab labs has shown its ingenuity during the covid-19 crisis with young students helping build protective medical equipment in collaboration with the local start-up ecosystem.

NORMATIVE FRAMEWORK, RESPONSIBLE INNOVATON IN DEMOCRATIZED INNOVATION ECOSYSTEMS: Democratized ecosystems, which build on open source approaches to new technologies, have developed unprecedented active local innovation hubs with incentives to better tailor tech applications to real social issues. Yet, states or regions can truly deploy, implement and flourish with new technologies only if they promote a normative environment that enables technological transfer to the grassroots level, delivering valuable products and services not only to the wealthy and powerful but also to social and economic peripheries.[122]

[121] https://www.theweek.in/news/sci-tech/2020/01/24/Kerala-set-to-get-two-design-fabrication-labs.html

[122] Li D. and Pauwels E. 2018. "AI & Global Governance: AI for Mass Flourishing," UN Centre for Policy Research, 15 October. https://cpr.unu.edu/ai-global-governance-ai-for-mass-flourishing.html

Just like any automated AI and biotech labs today, democratized innovation ecosystems could be targeted to produce malicious uses of AI and converging technologies. Yet, to avoid regulatory sanctions, most community labs have agreed to collaborate with security experts and develop codes of conduct and ethos of responsible innovation to be shared and implemented between peers.[123] Such an effort gives us insights into the kind of responsible practices that could ratify knowledge sharing between entrepreneurs, engineers and security communities. The next step is to foster legitimacy for democratized innovation by supporting mentors and their students to document and share their data, evidence and ethical concerns in ongoing conversations with regulators and society at large.

Because they function as a peer-review culture, community labs constitute an ideal ecosystem for mentorship in the most current engineering techniques and their related risk-benefit trade-offs. By the same token, these labs are the perfect place to start a continuing dialogue about how to adapt our regulatory standards to an increasingly democratized form of innovation. Community labs could become space where to perform socio-technical system analysis relevant to local needs and ethical expectations: mentors and engineers would guide students to think through all the possible effects of AI applications on their community. They would also engage with social impacts at every stage — conception, design, deployment and oversight.

Adaptive regulatory frameworks could be the starting point of a discussion to ensure safe and responsible citizen participation in AI and converging technologies. The way forward is to create a dialogue through which regulators can help citizens embed tailored oversight mechanisms into their endeavours. A more equal and inclusive world will not appear by chance. It will rest on the empowerment of those who can imagine globally beneficial intelligent designs.

[123] Kuiken T., 2016. “Governance: Learn from DIY biologists”. Nature. https://www.nature.com/news/governance-learn-from-diy-biologists-1.19507

SECTION 4 – CONVERGING RISKS & ETHICAL CONSIDERATIONS FOR HUMAN CAPITAL

Converging technologies have powerful implications for how the South Asia region will emerge from and be resilient to the current covid-19 crisis and future global security threats. Yet, they also impact the methods and practices harnessed by the World Bank Group and partners involved in improving and accelerating human capital outcomes. Three strategic trends are emerging that need to be understood to uncover the ethical implications of integrating converging technologies into human capital programs/projects:

● **First, private sector actors at the forefront of the digital economy are increasingly using the capacity of converging technologies for automated and predictive behavioural analysis**.[124] Powerful technological platforms, data-analytics firms and companies in the political, biometrics and security intelligence sectors, collect and analyse massive amounts of personal and behavioural data about populations to predict future needs, demands and lucrative markets.[125] These corporate actors, with their substantial competitive advantage in AI and predictive analysis, are interested in partnering with the human development sector to help expand the scope of digital identity and human security management systems.[126] In particular, the covid-19 pandemic crisis has shown the extent to which populations' digital identities and behavioural data can support epidemiological surveillance.[127]

● Second, the World Bank and other actors in the humanitarian sectors have started relying on the technological capabilities of digital platforms and private sector leaders in the field of AI, predictive data-analytics and biometric identity management systems.[128] Analysing the emerging ethical implications generated by such partnerships can help human capital experts anticipate unintended consequences and reflect on their own approach to normative leadership and risk management. **There is an urgent need to develop a "theory of no-harm"[129] when integrating converging technologies into strategies and programs aimed at accelerating human capital.**

● The two above trends indicate that World Bank' s programs, in their effort to improve human capital, will face increasing digital interdependence and heavy reliance on the complex private sector supply chains that produce converging technologies. The Covid-19 pandemic is also acting as a pressure to transfer human capital efforts "online," in cyberspace where governance rules have yet to be defined and strengthened. **The World Bank and its partners will therefore be confronted with heightened converging cybersecurity risks.**

In the aftermath of the covid-19 crisis, the World Bank, in its effort to improve and accelerate human capital, is about to face an upheaval, confronted with renewed questions about its capacity for building resilience, reducing inequality and supporting local communities' empowerment in times of

---

[124] Dance G., LaForgia M., and Confessore N., 2018. "As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants.". *The New York Times.* https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html; Fowler G., 2019. "How we survive the surveillance apocalypse." *The Washington Post*. https://www.washingtonpost.com/technology/2019/12/31/how-we-survive-surveillance-apocalypse/

[125] Zuboff S., 2020. "You are now remotely controlled." *The New York Times.* https://www.nytimes.com/2020/01/24/opinion/sunday/surveillance-capitalism.html

[126] Older M., 2019. "Google is Coming for Your Face.". *The Nation*. https://www.thenation.com/article/archive/immigrant-dna-data/; The Engine Room, 2018. "Biometrics in the Humanitarian Sector." *Oxfam*. https://www.theengineroom.org/wp-content/uploads/2018/03/Engine-Room-Oxfam-Biometrics-Review.pdf; Kinstler L., 2019. "Big tech firms are racing to track climate refugees." *MIT Technology Review*. https://www.technologyreview.com/2019/05/17/103059/big-tech-firms-are-racing-to-track-climate-refugees/

[127] Singer N. and Sang-Hun C., 2020. "As Coronavirus Surveillance Escalates, Personal Privacy Plummets." *The New York Times.* https://www.nytimes.com/2020/03/23/technology/coronavirus-surveillance-tracking-privacy.html

[128] Madianou M., 2019. "Techno-colonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises." *Social Media + Society.* https://journals.sagepub.com/doi/full/10.1177/2056305119863146; Kendall M., 2016. "Palantir using big data to solve big humanitarian crises". *The Mercury News.* https://www.mercurynews.com/2016/10/04/palantir-using-big-data-to-solve-big-humanitarian-crises/

[129] For early reflections on principles of due diligence and "Do no harm" in harnessing big data for conflict prevention, See Mancini F., 2013. "New Technology and the Prevention of Violence and Conflict." *International Peace Institute.* https://www.ipinst.org/images/pdfs/IPI_Epub-New_Technology-final.pdf p.24.

emergency and global public health crisis.[130] Beyond these challenges, the World Bank and its partners will have to learn how to anticipate and mitigate increasing information and cybersecurity threats.[131] **In this pressing context, World Bank actors will face a new array of temptations and opportunities to integrate cutting-edge technologies into their work, in particular predictive and automated behavioural analysis. Automated predictive algorithms are already used by corporations for securing cyber-networks, monitoring fraud in social protection, adapting education curricula to online personalised systems, and diagnosing an array of health problems, including stunting.** Without adequate foresight, risk assessment and normative leadership, the field of human capital may gradually rely on new, enhanced forms of behavioural monitoring/surveillance driven by technologies fully or partially made by private sector actors in still weakly regulated[132] supply chains.

COMMODIFICATION OF CHILDREN'S BEHAVIOURAL AND EMOTIONAL DATA – PROFILING, SURVEILLANCE & EMOTION MANIPULATION

Two of our case-studies – in education (emotional learning/Affectiva) and health (stunting/Child Growth Monitor) respectively – rely on the pervasive video capture of children's biometrics and emotional data. Yet, there is an urgent need to make clear that inferences about children's and students' behavioural and emotional data are subsequently used to train the neural networks owned by Ed Tech providers and larger legacy companies such as Intel and Microsoft. This is significant because both well-known and lesser-known companies use facial/biometrics and behavioural data-capture for applications beyond education. For instance, Affectiva, which develops AI technology for adaptive learning applications, including improving social and emotional skills, has already sold its technology to HireVue, another company that offers to screen job candidates based on how often they smile and whether they exhibit qualities like strength of character or social confidence.[133] Not only it is increasingly difficult to monitor how predictive emotional and behavioural analysis is used and potentially misused through complex supply chains, there is also a growing ecosystem of companies thirsty for children's data to improve their algorithmic services in other business contexts. We face emerging risks to see children's emotions and behaviours commodify to serve business and strategic contexts for which the class-room data was not intended (such as employment, retail, advertising, media and entertainment).

Such problem of misalignment between "aspirational" use and potential for misuse goes beyond the commercial and employment sectors, with implications for human rights, agency and human security. For instance, predictive behavioural and emotional analysis already enables hyper-personalized campaigns in which key demographics, including young voters, are manipulated to affect voting behaviours at crucial times.[134]

In countries where privacy and data protection are not translated into robust accountable mechanisms, domestic political parties can exploit sensitive population datasets and social media networks for spreading targeted propaganda, hate speech, mis and disinformation. In 2013 and 2017 Kenyan elections, divisive and inflammatory online propaganda, including graphic video violence, targeted ethnic and socio-economic population subgroups through mobile phone and social media networks as well as traditional media.[135] Each election witnessed more refined and precise strategies

---

[130] Mancini F., 2013. "New Technology and the Prevention of Violence and Conflict." *International Peace Institute.* https://www.ipinst.org/images/pdfs/IPI_Epub-New_Technology-final.pdf

[131] Pauwels E., 2019. "The New Geopolitics of Converging Risks." *United Nations University Press.* https://collections.unu.edu/eserv/UNU:7308/PauwelsAIGeopolitics.pdf

[132] Crawford K, Dobbe R., Dryer T., Fried G., Green B., Kaziunas E., Kak A., Mathur V., McElroy E., Sánchez A., Raji D., Rankin J., Richardson R., Schultz J., West S., and Whittaker M., 2019. "AI Now 2019 Report". *AI Now Institute.* https://ainowinstitute.org/AI_Now_2019_Report.pdf pp 19-22

[133] Idem

[134] Hern A., 2018. "Cambridge Analytica: how did it turn clicks into votes?" *The Guardian*. https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie

[135] Human Rights Watch, 2018b. "Kenya- Events of 2017". HRW. https://www.hrw.org/world-report/2018/country-chapters/kenya

for controlling spheres of information and exploiting political and emotional engineering targeted at segmented communities of Kenya's population, including its youth. Such strategies were crafted with the support of foreign data-analytics companies, Cambridge Analytica and the SCL Group, for profiling and influencing voters' behaviours.[136] Major political parties, the ruling Jubilee party and the opposing political group called the National Super Alliance (NASA), had also built and deployed widespread communication architecture to target young voters among the Kenyan population.[137]

In the near-future, malicious actors could rely on predictive behavioural analysis to identify the emotional triggers that push subgroups to violence, amplifying social engineering, psychological manipulation and other techniques of subversion and deception.

In another context, in cybercrime, the combination of emotional engineering with personal datasets can help craft even more convincing attacks that can hardly be recognised as a threat. Today, AI malware can watch, track and evaluate individuals' emotions, language and behaviour, impersonating trusted contacts within professional and personal social networks. Tailored communication generated by AI malware will therefore be almost impossible to distinguish from human peers' communications. An AI system that has been taught to study the behaviour of social network users and implement finely-targeted, personalized spear-phishing attacks on them, was able to perform more than 6 times as efficiently as humans and with a higher conversion rate.[138]

Accessing biometrics and behavioural children's data may also play an increasing role in cyber-bullying, with the capacity to use forgeries, face and emotion-datasets in the online pornographic industry (what is commonly called "Deepfake Porn").

In August 2019, researchers in Israel published a new method for making Deepfakes by creating realistic face-swapped videos in real-time, with no extensive facial data-training. Deep-learning algorithms – called FSGAN – can pinpoint facial biometrics features in a video, then align the source face to the target's face.[139] Algorithms that do not need to be trained on each new face target provide a powerful toolkit to create realistic video forgeries at scale and with minimal know-how. In their article, the researchers warn about the potential for democratizing video forgeries: "Our method eliminates laborious, subject specific data collection and model training, making face swapping and re-enactment accessible to non-experts."[140] The deployment of AI-enabled forgery technology will drastically alter relationship to evidence, truth and trust in online environments with unforeseen implications for vulnerable groups, in particular children and young women, at community and household levels. The capacity of a range of actors to influence public opinion with misleading simulations could have powerful long-term implications for social cohesion, with rising potential of discrimination and exclusion. In India's West Bengal region, Rohingya refugees, who fled exactions in Myanmar, are now demonized in violent speech that rapidly metastasize on WhatsApp.[141]

---

[136] Privacy International, 2018b. ''Further questions on Cambridge Analytica's involvement in the 2017 Kenyan Elections and Privacy International's investigations''. *Privacy International*. https://privacyinternational.org/long-read/1708/further-questions-cambridge-analyticas-involvement-2017-kenyan-elections-and-privacy

[137] Muthuri R., Monyango F., and Karanja W., 2018. "Biometric technology, elections, and privacy: Investigating privacy implications of biometric voter registration in Kenya's 2017 Election Process." Centre for Intellectual Property and Information Technology Law. https://www.cipit.org/images/downloads/CIPIT-Elections-and-Biometrics-Report.pdf

[138] Seymour J. and Tully P. "Weaponizing Data Science for Social Engineering." https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-SocialEngineering-Automated-E2E-Spear-Phishing-On-Twitter.pdf

[139] Cole S., 2019. "This Program Makes It Even Easier to Make Deepfakes". *Vice News.* https://www.vice.com/en_us/article/kz4amx/fsgan-program-makes-it-even-easier-to-make-deepfakes; ODSC - Open Data Science, 2019. "FSGAN: Subject Agnostic Face Swapping and Reenactment," *Medium.* https://medium.com/@ODSC/fsgan-subject-agnostic-face-swapping-and-reenactment-2f033b0ea83c

[140] Nirkin Y., Keller Y., and Hassner T., 2019. "FSGAN: Subject Agnostic Face Swapping and Reenactment" *arXiv*. https://arxiv.org/pdf/1908.05932.pdf

[141] Goel V. and Rahman, SA., 2019. "When Rohingya Refugees Fled to India, Hate on Facebook Followed ". The *New York Times.* https://www.nytimes.com/2019/06/14/technology/facebook-hate-speech-rohingya-india.html

HUMAN RIGHTS IMPLICATIONS OF BEHAVIOURAL SURVEILLANCE

The use of systemic, automated and predictive behavioural analysis in the field of human development, potentially for improving human capital, is likely to pose new and fundamental challenges for protecting human rights in fragility contexts. In this era where AI combines with powerful data-capture technologies, such as biometrics, facial and emotion-recognition, algorithmic surveillance amplifies what Michel Foucault called "biopolitics," a series of interventions to regulate society's collective body.[142] There is a current tendency to underestimate how converging technologies can be designed to anticipate and influence human behaviours for social and political control, with corrosive human rights implications. These implications include limits to self-determination and political agency, lack of privacy and sensitive data-protection, exposure to pervasive data-security breaches, and new forms of censorship in the virtual civic space.

Accountability and Remedy: Any autonomous data-capture and algorithmic system that affects human well-being and agency is a test to the norms of remedy and accountability. Audit studies have shown that AI systems can act in unpredictable ways.[143] If such a system is diverted from its initial purpose and misused in a human rights harm, it could be difficult to ascertain responsibility. Responsibility might be fragmented between the designer of the system, the supplier of the system, a third party (such as a vendor) or the international organisation sponsoring the use of AI in human development. Some experts believe that using AI systems will create an "accountability gap", making it difficult, if not impossible, for harmed parties to obtain access to remedy and fair treatment by a judicial system.[144] There is also considerable concern that AI technologies could weaken the international rule of law, especially when employed in intelligence and surveillance activities, by facilitating extrajudicial actions like lethal drone strikes.[145]

Privacy: Converging technologies pose profound challenges to privacy, because they can automate the detection of "anomalies" or "abnormal behaviour" in very large amounts of bulk, routine data about individuals and crowds. For example, image and speech recognition algorithms can detect objects in blurry photographs or separate voices in crowded environments. The dual-use potential of converging technologies exacerbates the problem, as technologies built for lawful uses can easily be adapted to facilitate expanded surveillance in violation of human rights principles.

Right to Self-Determination: By introducing new opportunities for authoritarian states or non-state-violent-actors to control populations, AI threatens political participation and freedom of movement. The same way, AI-led surveillance technologies also pose a real threat to peaceful assembly and protest.

Non-Discrimination and Minority Rights: Current facial recognition algorithms, in their optimization processes, fail to discern the features of darker skinned faces with the same rates of accuracy as they detect the geometry of lighter faces.[146] Minorities could be stigmatized and ostracized in new powerful ways. A major concern when it comes surveillance and minority rights is the potential for automated ethnic profiling. In recent years, multiple investigations by human rights defenders have unveiled how China's government imposed on the Uighur population facial-recognition tracking and the collection of biometric data, including DNA samples and voice samples.[147] With time, Chinese authorities have

---

[142] Foucault M. 1976. "The History of Sexuality. An Introduction, Vol. 1." *Vintage Books*. p. 138-139.
[143] Yampolskiy R., 2019. "Unpredictability of AI". https://arxiv.org/ftp/arxiv/papers/1905/1905.13053.pdf
[144] Courtland R., 2018. "Bias detectives: the researchers striving to make algorithms fair ". *Nature.* https://www.nature.com/articles/d41586-018-05469-3
[145] ICRC, 2019.
[146] Buolamwini J. and Gebru T., 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". *Proceedings of Machine Learning Research*. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
[147] Human Human Rights, 2018c. "Eradicating Ideological Viruses China's Campaign of Repression Against Xinjiang's Muslims". https://www.hrw.org/report/2018/09/09/eradicating-ideological-viruses/chinas-campaign-repression-against-xinjiangs ; O'Brien D., 2019. "Massive Database Leak Gives Us a Window into China's Digital Surveillance State". *EFF*. https://www.eff.org/fr/deeplinks/2019/03/massive-database-leak-gives-us-window-chinas-digital-surveillance-state

now deployed a vast system of facial recognition algorithms that have been trained to associate certain skin tones and facial features with the concept of Uighur ethnicity.[148] This form of profiling makes China a leader in applying AI to monitor subpopulations, with the potential to export a new type of automated racial surveillance.

For instance, Thailand and Vietnam have already adopted a similar approach to the Great Firewall – relying on a combination of legislative and technological tools to regulate the internet domestically. During the covid-19 crisis, Vietnam has relied on massive social closures and extensive surveillance of citizens.[149] From the early onset of the pandemic, Vietnam shuttered non-essential businesses and schools and enacted large-scale quarantines—tens of thousands of citizens have been placed in "quarantine camps" run by the military. Vietnam's aggressive monitoring and surveillance of citizens has been supported by the government's large network of informants, which has helped to identify and quarantine those suspected of infection and those who have been in contact with them.

CIVILIAN DATA AND INFORMATION INFRASTRUCTURE SECURITY

As the field of human development and human capital integrates AI and data-capture technologies, the World Bank and its partners will be confronted with heightened converging cybersecurity risks. In the aftermath of the Covid-19 crisis, an array of processes for population monitoring and tracking have become digital, augmented by AI and sensing technologies. Under the same impulse, data-optimization and converging technologies, operating in cyberspace, have the potential to modernize and impact human development and human capital.

In the very near-future, the World Bank and its partners will need to thoroughly secure the behavioural and contextual information they collect about populations. Protecting such sensitive datasets from digital manipulation and cyber-exfiltration will be a complex challenge but crucial to ensure the security of civilians and critical information infrastructures.

For instance, remote sensing imagery, from satellites or drones, can provide information about refugees' settlements, populations' relatively precise locations, human rights violations and structural damage in disaster relief. A cyberattack by a state or non-state-violent actor could aim at exfiltrating such sensitive information for surgical offensives or strikes.

The fundamental nature of cyber-attacks is also about to change dramatically in coming years. Previous cyber operations have largely been about access – i.e. getting or preventing access to sensitive data. Yet, more sophisticated operations are already targeting data and algorithmic integrity.[150] Autonomous malware can learn to surreptitiously poison datasets or manipulate algorithmic models. This new type of adversarial attacks[151] could alter the integrity or function of hate speech monitoring platform, early-warning systems and AI-powered analytics tools. The ultimate harm would be to corrupt intelligence collection and analysis, or allow attackers to control a predictive system.

Technically, this prospect is not far-off. In 2019, computer scientists in Israel used algorithms to hack hospital CT scans, producing false, non-existing tumours that conform to a patient unique anatomy.[152] The experiment led to 90% of misdiagnosis. The entire attack can be fully automated so that once the malware is launched into a hospital network, it will operate on its own, to find and alter scans, even searching for a specific patient's name. A group at Harvard University has also tested adversarial

[148] Mozur P., 2019. "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority". The New York Times. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html
[149] https://www.csis.org/analysis/strengths-and-vulnerabilities-southeast-asias-response-covid-19-pandemic
[150] Pauwels E., 2019b
[151] Muppidi S., 2018. "Adversarial AI: As New Attack Vector Opens, Researchers Aim to Defend Against It ". *Security Intelligence.* https://securityintelligence.com/adversarial-ai-as-new-attack-vector-opens-researchers-aim-to-defend-against-it/
[152] Mirsky Y., 2019. "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning." https://arxiv.org/abs/1901.03597

attacks against algorithms used to diagnose skin cancer images, showing that such attacks need only modifying a few pixels in the biopsy picture to modify the final diagnosis.[153]

Similar attacks where data from biometrics ID or face datasets is exploited have led to new cyber-threats. This is what the author call "precision biometric attacks," which are expanding with algorithms learning to impersonate someone's expressions in a video or audio file. Forgeries or "Deepfakes" could be used in incredibly convincing spear phishing attacks that civilians or mediators would have a very hard time to identify as false. Already, impersonation attacks are on the rise: about two-thirds of businesses saw an increase in forgeries in the last 12 months.[154] In March 2019, cyber criminals relied on machine learning voice spoofing to commit a cybercrime by reproducing the voice of a CEO, demanding a fake transfer of about $240,000.[155] In 2018, IBM detected an AI malware that can hide a cyberthreat, such as WannaCry, in a video conference application, and launch only when it identifies the face of the target.[156] In the context of human development and humanitarian assistance, the vulnerability of biometrics databases is a constant and long-term concern as several cyberattacks have already compromised the biometrics and personal data of vulnerable populations.

LOGIC OF EXPERIMENTAION, TECHNICAL AND PREDICTIVE FAILURES

Converging technologies may improve the ability to automate a range of services crucial to improving human capital and monitor remotely the needs of vulnerable populations, particularly in time of global pandemics. Yet, automated remote management may give the World Bank and its partners a "false sense of informed decision-making"[157] and prevent them to assess if algorithmic monitoring performs with accurate predictive value. Meeting the ethical expectations and needs of local populations, such as ensuring meaningful accountability, might be undermined in the process.

There are significant ethical considerations centred around the harm that could be generated by technical problems and failures in predictive value.[158] The limits to using AI and data-capture technologies for predictive behavioural analysis in human capital sectors are significant: the lack of accurate, updated and representative datasets; the quality of data-curation and algorithmic training; cognitive, gender, racial, historical or economic biases; and a dearth of theoretical and statistical knowledge about health/diseases and developmental/learning/cognitive phenomena critical to human capital.[159] Actors on the ground and those taking decisions remotely will need to understand the computational techniques used in specific AI and converging technological applications. They will also need to understand the datasets in use, particularly how data is collected and what biases those datasets may represent. In general, such an effort requires to examine how data and analytical processes are embedded within the socio-technical infrastructure that surround vulnerable and underserved populations.

For instance, in the field of affective computing, there remains little to no evidence that new affect-recognition applications have any scientific validity. In February 2019, researchers at Berkeley found that in order to detect emotions with accuracy and high agreement requires context beyond the face

[153] Finlayson S., Bowers J., Ito J., Zittrain J., Beam A., and Kohane I., 2019. "Adversarial attacks on medical machine learning". *Science.* https://science.sciencemag.org/content/363/6433/1287.full?ijkey=OXnSsEp.Iagl6&keytype=ref&siteid=sci

[154] Darktrace, 2018. "The Next Paradigm Shift AI-Driven Cyber-Attacks." https://pdfs.semanticscholar.org/6b18/6268d00e891f3ed282544ac5833c01a2891c.pdf p 5

[155] Stupp C., 2019. "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case". *The Wall Street Journal.* https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

[156] Osborne C., 2018. "Deeplocker: When malware turns artificial intelligence into a weapon." ZDNet. https://www.zdnet.com/article/deeplocker-when-malware-turns-artificial-intelligence-into-a-weapon/ ; Kirat D. et al, 2018. "DeepLocker: Concealing Targeted Attacks with AI Locksmithing." Black Hat USA 2018.

[157] United Nations, 2019. p 18

[158] Bazzi S., Blair R., Blattman C., Dube O., Gudgeon M., and Peck R., 2019. "The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia". *NBER*. https://www.nber.org/papers/w25980.pdf

[159] Guo W., Gleditsch K., and Wilson A., 2018.

and body.[160] Given the high-stakes of using affect-recognition systems to measure children's performance and potentially, youth's employment prospect, the scientific validity of those applications is an area in particular need of research and policy attention. The problem of "automation bias" – humans tend to stop questioning suggestions from automated decision-making systems and to ignore contradictory information – significantly raises the stakes for the use of AI emotion-analysis in education.[161]

In the foreseeable future, AI technologies will increasingly contribute to analyzing mass routine data and extracting predictions about populations' behaviors and performance, including in human capital, for higher-level analysis by human operators. Yet, the amount of bulk data in monitoring operations means that these systems will be operating with a high degree of autonomy. Automated data-optimization and assessment/diagnosis, in health, education and social protection, might become the norm. For instance, in the context of the covid-19 crisis, manual measurements and physical exams of children, suspected of suffering from stunting, have been suspended because it became impossible to keep a safe distance. Automation may lead to reducing the size of and the support to the field workforce that usually conducts health visits with unintended, harmful consequences. One of these consequences could be a lack of training of field workers about what healthy morphological and cognitive development means, how it materializes in a healthy child, and what factors account for creating positive child's growth conditions. Less trained field workers may prevent the needed, expected behavioral change by members of the household when it comes to nutrition, health and cognitive development. What are the larger unintended – or un-anticipated – implications at the household and community levels? Similar arguments can be made in education, where AI-led personalized learning system might impact the role of teachers, the funding allocated to their jobs, and the teacher-student relationship.

This paper does not argue for impugning the entire field of affective computing or converging technologies. But it attracts attention to the often invisible or forgotten business dynamics (data commodification) that require converging technologies to become part of a logic of solutionism, reductionism and experimentation on vulnerable populations. This paper also stresses how most normative frameworks are not future-proof, not resilient to future contexts where converging technologies – interacting with complex social systems – create potential for emerging threats, as well as ethical and governance failures. Often, our normative frameworks are constrained by legacy approaches and blind to the unforeseen ways through which tacit knowledge (math/algorithms), data and automated technologies can be decentralized for dual-use.

[160] Zhimin Chen and David Whitney, "Tracking the Affective State of Unseen Persons," Proceedings of the National Academy of Sciences, February 5, 2019, https://www.pnas.org/content/pnas/early/2019/02/26/1812250116.full.pdf.

[161] Cummings M. "Automation Bias in Intelligent Time Critical Decision Support Systems". American Institute of Aeronautics and Astronautics. https://web.archive.org/web/20141101113133/http:/web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf

SECTION 5 – RECOMMENDATIONS TO THE MULTILATERAL SYSTEM AND INTERNATIONAL COMMUNITY

Technological disasters with irreversible environmental impacts provide a strong cautionary tale for how the multilateral system should shape its human capital engagement with converging technologies. Statistics tell one common, troubling story about environmental disasters: the poor and vulnerable suffer more than the rich. A similar calculus applies to the erosion of human capital with children, women, migrants, refugees and other minorities suffering corrosive long-term disempowerment.

Two opposing stories involving tech and environmental harms should help the World Bank reflect on how to build and deploy a "Theory of No-Harm" in the field of human capital.

REVERSING POWER & KNOWELDGE ASYMMETRIES – TWO STORIES:

BHOPAL & THE SILENT VALLEY

Bhopal is known as one of the worst industrial disasters in history. In December 1984, a massive amount of water unexpectedly inundated the underground storage tanks containing methyl isocyanate at the Indian subsidiary of the American chemical company, Union Carbide Corporation.

The brisk wind that night picked large clouds of toxic gas from the leaking tanks to and across Bhopal, a working-class city, with its vast and crowed slums, south of New Delhi. The gas settled at low levels where poor families – many from the city's poorer Muslim quarter – were breathing and sleeping on the ground. In an hour, the gas had submerged the frail homes of ten thousands of people, killing hundreds in their sleep, annihilating whole families, provoking terrible injuries and illnesses to those who survived the tragedy. Many questions went unanswered: How did the leak occur? Who was responsible? Could it have been prevented? How much was known in advance, or not known, about methyl isocyanate's (MIC) long and short-term effects on human health? Why was a substance of such toxicity stored onsite in a dense urban setting? What emergency measures were in place and why were there so inadequate? Whose responsibility was it to care for the victims, not just in the accident's immediate aftermath but through long-term medical and social monitoring of chronic illnesses?

These questions – and subsequent chances for accountability and redress – were never addressed in public forum as the company continued to claim the disaster happened as the result of criminal intent, not an accident. The irreversible damages suffered by Bhopal most vulnerable populations were difficult to ascertain and prove in the law. Chemical exposure cases constitute complex problems of evidence and proof because long-term negative effect such as cancers, chronic diseases, trauma and mental illnesses, often appear long after exposure, and because precise toxicity studies involving MIC had been neglected. When high-quality evidence is difficult to establish, methods for how to collect adequate data are not easily available, disagreement and misinformation abound. Lines of responsibility become fragmented along large, complex supply chains.

As explained in our case-studies, similar unforeseen threats could endanger, erode human capital if safeguards are not in place to protect citizens, in particular the most vulnerable, children, women and minorities. Responsibility for exploiting and monetizing behavioural, emotional and biological data about children, could be increasingly fragmented along the complex supply chains of AI and converging technologies. Irreversible damage could harm the next-generations if, long-term, human rights impact assessment and data protection are not mandatory, operational and effective.

The Save Silent Valley Movement tells another story, one where poets, activists, citizens and scientists shared ecological and environmental knowledge, learned about concrete outcomes, and organized into a large, powerful social movement to save Kerala's oldest evergreen forests and biodiversity hotspot from being submerged and irreversibly damaged. The movement was opposing the construction by the Indian government of an enormous dam project with financial support from the World Bank.

The Save Silent Valley Movement eventually led not only to the suspension of the dam's project, but became a landmark moment for environmentalists in India. Many other social and environmental forms of resistance came to birth, including against the Narmada Bachao Andolan project. The Narmada River project had planned to build a massive dam – The Sardar Sarovar dam – which would have displaced large numbers of tribal populations fleeing rising waters. The World Bank had chosen to finance a portion of the Narmada River Valley project in 1985 without the benefit of a full environmental impact study. Concerns raised by environmental groups forced the Bank to undertake its first independent review of a bank-financed development project.

A 1989 World Bank paper - before strict environmental standards were adopted – mentions the role of both, the Save Silent Valley and Narmada River movements, among others, and states: "The Bank recognizes that some environmental effects may require many years before they become evident. Consequently, Bank policy requires consideration of the environmental aspects of projects in a longer time frame (i.e., decades and longer) than is the case for most other projects. Since considerable uncertainty exists about the magnitudes of some long-term effects, the Bank emphasizes the importance of prudence when assessing environmental impacts, especially when these are potentially irreversible, as in the case of extinction of species or conversion of ecosystems." India's powerful resistance movements contributed to force the Bank to adopt new environmental and social standards and suspend investment in large dams.

Both stories, in their differences, show how we need a revolutionary response to emerging complex and uncertain situations demanding drastically new ways of thought and a self-conscious departure from post-damage corporate or government liability doctrines.

When it comes to human capital projects, there is an urgent need to consider if there could be irreversible damage to empowerment and agency, especially amongst children, women and marginalized populations. Like in the context of environmental engineering, such potential for "dual-use" and long-term harm is difficult to assess, but this is an even more critical reason for the Bank to define a "Theory of No Harm," along a theory of change in human capital.

As we have been facing increasing threats to our ecosystems on our burdened planet, we are already confronted with signs of disempowerment and erosion of human capital. More sobering, these threats to human capital have the potential to create more powerful divides, digital underclasses, with corrosive impacts for young women, minorities, climate and conflict-migrants and other vulnerable populations. Two case-studies in our paper – AI for personalised learning and for diagnosing stunting – illustrate this new species of risks to civilian populations, including children. There could be irreversible damages to human capital across South Asia.

In the case of AI emotional analysis for personalised learning, potential violations to child rights are salient. There are (1) serious questions about effectiveness, validity and representativeness of training data; (2) financial incentives and the wellbeing of school-children do not necessarily align in this situation where pervasive commodification of children's data is a gold mine for the private sector; (3) mining the emotional lives of children infringes on a set of child rights[162], especially when the value extraction does not always serve the wellbeing of those children; (4) it is problematic to use inferences about children's emotions to train neural networks deployed for other commercial purposes (such as advertising); (5) lack of proportionality and breach of data-purpose rules where in-class data may be used for other socially determining purposes (such as social scoring); (6) the short and long-term impact of children suffering of chilling effects in the classroom; and lastly, (7) existence of data minimisation principles that should ask if emotional AI is necessary for successful education. In the case of emotional AI and facial coding in the classroom, there is therefore a potential for increasing misalignment with current and near future social values.

[162] the child's right to freedom of thought (Art. 14) and privacy (Art. 16), the right to develop full potential (§1 Art. 29), the child's right to liberty (§2 Art. 29), and the child's right to be protected from economic exploitation (Art. 32)

In light of the accelerating pace of datafication and tech convergence, there is an urgent need for countervailing forces to the logics of solutionism and surveillance capitalism that are imported by large corporations into the humanitarian sector. Without safeguards, empowerment and accountability mechanisms, the World Bank will increasingly find itself part of a large attempt at commodifying personal, behavioural, emotional and biological data about populations, in a technological revolution whose unforeseen consequences have not been anticipated or disclosed by private and public sector actors.

□ In addition to a theory of change, the World Bank needs a "Theory of No-Harm," which means a form of normative foresight to anticipate and weigh the human capital benefits of converging technologies against the costs to fundamental human rights.

In the coming years, the Bank is about to face one of its most difficult strategic and ethical choices: whether it should integrate, in its mandate, converging technologies that are inherently designed for predictive analysis based on mass populations' behavioural data. Without adequate foresight, risk assessment and normative leadership, World Bank projects may progressively rely on new, enhanced forms of behavioural surveillance driven by technologies fully or partially made by private sector corporations. Such transformational shift begs significant questions about the Bank's normative methods, ethos and need for safeguards:

For an international institution that (ostensibly) operates in the universal public interest, what does it mean to integrate forms of automated and predictive behavioural surveillance in human capital programs? What normative vision and "Theory of No-Harm" does the Bank need to develop as it adapts to the changing nature of human development and human capital under technological convergence?

To date, the human development and the humanitarian sectors – have not yet fully developed and operationalised a common "Theory of No-Harm." A "Theory of No-Harm" would consist in developing adequate epistemic and normative methods to weigh the benefits of harnessing converging technologies, in particular automated and predictive behavioural monitoring, against the costs to civilian security and human rights. Within the Bank, there are no agreed and stress-tested methods to assess the ethical, security and human rights implications of delegating some strategic elements of building, deploying and accelerating human capital to automated predictive technologies. Therefore, a substantial accountability gap exists when no methods and cross-sector collaborations have been conceived and deployed to anticipate unforeseen misuses and long-term impacts of AI and data-capture technologies on vulnerable populations. Three broad recommendations emerge on devising a "Theory of No-Harm," its epistemic methods and collaborations cross-sectors:

● NORMATIVE METHODS TO DEVELOP A "THEORY OF NO-HARM" → NORMATIVE & INCLUSIVE FORESIGHT

Improving the human capital agenda will require normative and inclusive foresight to anticipate the nature and scope of interdependent societal and (cyber)security risks that may threaten vulnerable populations. Such foresight will become critical as converging technologies become more accessible to a wider range of state, non-state and corporate actors in weakly regulated supply chains.

Foresight can play a normative role and support a "Theory of No-Harm" by helping the Bank and its partners envision a range of normative & policy scenarios on how to manage the tension between 1) the need to build, deploy, accelerate human capital outcomes, bridging the digital, cognitive and health gaps, and 2) the imperative to prevent, minimize, mitigate and account for civilian and human rights harms. As an opportunity strategy, foresight methodologies can help the Bank and experts on the ground leverage ethical and normative solutions. As a form of interdependent risks management, these methods can help provide feedback loops to prevent or mitigate security, ethical and governance failures across systems and sectors. For instance, human capital experts within the Bank could work on a set of normative issues such as:

1. **Populations' Data Privacy & Security**: How to manage the powerful tension between the increasing opportunity to apply algorithmic analytics to mass populations' data (think of facial recognition, drone-driven monitoring, sentiment analysis) versus the privacy principles of data minimization and proportionality in data collection? To what extent have these privacy principles been thoroughly adapted to the contextual issues in education, health, nutrition and social protection within human capital? Under what time limitation and specific purpose will personal and behavioural data about populations and individuals (informants) be stored? Who (corporate actors, governments, civil society, citizens) will have access to these sensitive datasets and what types of security measures will be taken at all levels to ensure the integrity and safety of the data?

2. **Accountability**: Converging technologies, in particular new forms of predictive behavioural surveillance, could seriously disrupt the lives of vulnerable populations. In context of frail or inexistent data-protection and redress mechanisms, disproportionate access to populations' data by state and private sector actors may create new power asymmetries. How should the Bank anticipate and mitigate such power asymmetries and accountability gaps? Developing mechanisms for mitigating critical data incidents would support a common "Theory of No-Harm."

3. **Inclusion:** The current Covid-19 crisis has emphasized the need for rethinking inclusion in human capital efforts, by shifting to digital forms of engagements with populations, including women and youth. Yet, in education for instance, a significant tension exists between such needed form of digitization, but also the imperative to understand, anticipate and mitigate harmful forms of emotion manipulation, hate speech, cyber-bullying, child' s online exploitation, disinformation operations in online learning environment. As well explained by UNICEF, young "victims are not virtual."[163]

Normative and inclusive foresight should imperatively include civil society organisations as they have been at the forefront of the analysis and reporting of how behavioural surveillance may lead to human rights violations. One ultimate benefit of engaging local civil society organisations in normative foresight is not to deploy AI and converging technologies for human capital where we believe there are inadequate safeguards to protect human rights.

● ENGAGEMENT WITH GOVERMENTS & PRIVATE SECTOR – POLICY CO-CREATION IN PUBLIC-PRIVATE COLLABORATIVE EFFORTS

The World Bank needs to work with governments in client countries to help them better understand the interdependence between converging security challenges and ensure coherence among multiple policy and normative efforts spurred by states, private sector and civil society. Nowhere is this leadership and coherence more needed than at the convergence of dual-use technologies, human development and human capital, and nowhere does the Bank have more knowledge gaps.

There is also a crucial need to understand how to meaningfully collaborate with the global and local private sectors on those knowledge gaps. There is growing interest to drive developments in Ed Tech and health diagnostics technologies from both, small start-ups in behavioural data-analytics and large strategic intelligence corporations. Yet, the complex supply chains of converging technologies are still weakly regulated, which could lead to harms that would be unforeseen or unaccounted for through audits and redress mechanisms.

In 2018 and 2019, many companies produced AI ethics principles and due diligence statements.[164] Several shortcomings exist in this normative effort. First, the vast majority of these statements were

[163] https://www.unicef.org/rosa/sites/unicef.org.rosa/files/2018-03/Victims_are_not_virtual.pdf
[164] Pauwels E., 2019.

conceived by organizations whose leadership operate in the global North. Principle statements and the ethical priorities of the global South, let alone populations suffering from human capital erosion, are often absent from these AI normative maps. Second, obvious limits exist to self-regulation and corporate ethical principles. These principles need to be translated, materialized into viable normative practices that can be overseen and tested for transparency and accountability. Third, technological convergence and its particular implications are rarely understood when private sector actors define technical and normative standards. Responses to these three shortcomings have to be discussed by the World Bank Group and private sector actors under the umbrella of International Human Rights Law and the UN Guiding Principles of Business and Human Rights.

In this effort of normative translation and leadership, the Bank should incentivize engineers of AI and converging tech companies to share their unique visionary capacity and expertise to anticipate emerging threats in their domains. Their knowledge is instrumental when trying to delineate the drivers, nature, potential impact, likelihood and velocity of emerging technological risks. They should invest time and expertise in collaborating with the Bank to identify potential threats to human capital. Engineers and entrepreneurs can help policymakers discuss governance models that rely on "predictive accountability" for converging technologies. Under "predictive accountability," inventors and producers of dual-use technologies would conduct foresight to anticipate potential scenarios of technological misuses. Equipped with such predictive analysis, they would be able to improve technological design to account for and mitigate unintended consequences.

Human rights impact assessment: To be even more successful, policy co-creation efforts should include human security and human-rights impact assessment from the beginning, to assess converging technologies' risks and benefits, and define guiding principles for ethical and safe design. Human rights impact assessments provide a step-by-step evaluation of the impact of practices or technologies deployed in a given context on aspects such as privacy, data agency and self-determination. Collaborations should take place between private sector actors, external policy experts and World Bank teams that systematically monitor for ethical breaches and violations of human rights on the ground, impacting vulnerable populations. Building relationships between World Bank operational teams and human rights' labs inside AI companies (for instance, with Microsoft Human Rights' Lab) would constitute an interesting collaborative network to foster principles of predictive accountability.

In 2018, the UN Special Rapporteur for freedom of expression, David Kaye, has recommended that companies should account for discrimination at both the design and the deployment level of AI systems, and design AI-led tech systems that are non-discriminatory and account for diversity.[165] He has suggested that states and companies might be obligated to conduct human rights impact assessments and public consultations during the design and deployment of new AI systems, or existing systems in new markets.[166] He has also recommended that states should ensure that human rights are central to the design, deployment and implementation of AI systems. These recommendations offer concrete ways to ensure that states make an effort to prevent companies from violating human rights as they build and deploy AI. The recommendation about human rights impact assessments when technology is used in new markets is especially valuable for SAR countries since it acknowledges that it can be risky to import technology, designed by outsiders for societies of abundance, directly to SAR countries. It accounts for local context. AI systems need to be audited, assessed in situated realities.

Human rights organisations have also pointed out that discriminatory AI systems might violate the right to social security.[167] They may also affect states' obligation to ensure that people are able to

[165] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression to the UN General Assembly, Seventy-third session, 2018.
[166] Report of the Special Rapporteur (2018).
[167] Article 22 of the Universal Declaration of Human Rights as well as Article 9 of the International Covenant on Civil and Political Rights. See Human Rights Watch, Submission to the UN Special Rapporteur on extreme poverty and human rights, May 2019.

access the right to work,[168] necessitating efforts to enable people whose skills and jobs are affected by AI to acquire new skills and competence so that they are able to work.

As demonstrated with the cautionary tales of Bhopal and the Silent Valley in the environmental sector, there might be an increasing role for the World Bank, even in collaboration with civil society actors, to use its soft normative power to positively influence the adoption of impact assessment at different levels of design and deployment of technologies for human capital. Through another analogy, we have recently seen the impact that civil society actors and citizens within social movements of resistance can have over the use of facial recognition. In the wake of the 2020 spring protests in the US and globally, following continuous advocacy by civil society groups such as ACLU, the use of facial recognition in police body camera is now facing potential moratorium in the State of California, Oregon and New Hampshire.[169] While Microsoft[170] has long argued for regulatory frameworks around facial recognition, other large corporations like IBM and Amazon are reconsidering the use of such intrusive technology,[171] in particular selling such tech products to law enforcement.[172] But such window of reflection will only last if it is supported by soft power and normative leadership coming from legitimate institution serving the universal public interest like the World Bank and UN agencies.

● BETTING ON LOCAL COMMUNITY EMPOWERMENT IN AN AGE OF TECH CONVERGENCE

At program and project levels, World Bank teams need to go beyond performing normative foresight, human rights impact assessments and building robust policy and legal safeguards. There is also an urgent need to contribute to building the broader knowledge and citizen capacity, which is so critical to the "empowerment' angle of the human capital project. Most of South Asia's populations are in no position to anticipate for themselves either immediate human capital benefits or improved long-term prospects when technologies, designed elsewhere, often in societies of abundance, are directly imported and imposed on them without contextual thinking and adaptation (what refer in Section 3 and in the definition section in appendix as "socio-technical system analysis"). Section 2 and 4 explains the negative implications, for poor families in Kenya, of a private tech-driven education program (BIA) that is designed and imported by external corporate actors with for-profit incentives.

The hard truth is that inequality – not only of access, but even more of anticipation – emerges as a constant, unresolved limit to human capital empowerment of vulnerable populations in South Asia. The technological achievements (with their biases and unforeseen consequences) of wealthy societies are too often just transposed as the anticipatory horizons of the less privileged. This assumption is rarely defined in reverse: What technological futures would empower the large swaths of populations that live on less than two dollars a day? Answering such questions would require to rely on a trust-based and human-centred design approach to conceive technologies for the underserved.

Human capital responses to crises, when they include converging technologies, have to be designed, deployed and socialized on a trust-based approach, with the assurance that such responses will not, in the present and future, further erode human capital through automation of employment or surveillance. Some regions and states – the State of Kerala in India is a good example – have promoted inclusive innovation shaped to local needs and ethical expectations. Kerala demonstrates at the core of its model that social cohesion and resilience will matter more than technological solutionism.

The question remains though: how can we build the broader knowledge and citizen capacity, which is so critical to the "empowerment' angle of the human capital project?

---

[168] Article 6(1) of the International Covenant on Economic, Social and Cultural Rights.

[169] https://securitytoday.com/articles/2019/10/10/california-to-become-third-state-to-ban-facial-recognition-software-in-police-body-cameras.aspx

[170] https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/

[171] https://www.ibm.com/blogs/policy/facial-recognition-susset-racial-justice-reforms/

[172] https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/ ; https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition

Open Science, Open AI and Bridging Rising Knowledge Asymmetries: If we want to address society's most pressing and persistent challenges, from climate change to inequality, then technology will have a leading role to play. Scientific breakthroughs facilitated by AI could make the crucial difference, by helping to discover new knowledge, ideas and strategies in the areas that matter most to us all. But increasing public concern about lack of transparency and accountability within the technology industry should serve as a wake-up call. There are at least two epistemic asymmetries between the world of tech and the real world.

First, the disconnect between people who develop technologies and the communities who use them. Within the top AI labs, the employee base is unrepresentative when it comes to gender, race, class and more. Technology isn't value neutral, and it needs to be built and shaped by diverse communities if we are to minimise the risk of unintended harms. This is an urgent problem. Women and minority groups remain badly underrepresented, and leaders need to be proactive in breaking the mould.

Second, there's an asymmetry of information regarding how technology actually works. Solving this has to be a collaborative effort, and requires new types of organisations that facilitate deep understanding of how complex algorithms operate and their impact on society. This takes courage, trust and the prioritisation of real debate and engagement over the comfort of institutional roles, in which activists, governments and technologists are often more likely to criticise each other than to work together. It also requires more visibility into how data are used.

Mainstream discussions of AI assume that a small number of global tech firms will control the technologies that will affect the lives of massive populations. Recently, bitter criticisms have emerged of large private tech platforms that favour profits over public interest, undermining public trust. As a result, strategic discussions in AI governance circles have increasingly focused on defining ways to restrict and regulate the large corporations that lead the AI industry.

There is an alternative path for developing and deploying AI, what proponents of open innovation call Open Science, Open AI or "AI from the grassroots." In Kerala, India, democratized ecosystems, which build on open source approaches to new technologies, have developed surprisingly active local dynamism with incentives to better connect tech with real social issues. Increasingly, AI should be considered as a tool of what Edmund Phelps has coined "mass flourishing." Nations or regions can truly develop, implement and grow with new technologies only if they maintain a regulatory environment that enables technological transfer to the grassroots level, delivering valuable products and services not only to the wealthy and powerful but also to social and economic peripheries.

AI naturally lends itself to this kind of local, decentralized approach. Knowledge and means of technological production—the two key ingredients for innovation today—are available to anyone with internet access. Most AI technologies can be obtained via open source licenses. Many core learning algorithms are available on public platforms. For example, the means to build the kind of devices that run AI on smartphones, robots and self-driving cars are all available within the open innovation and production ecosystem.

Home-grown innovators around the world can now leverage global resources and tailor tech production to suit local needs, developing a knowledge advantage over their global competitors. As local entrepreneurs and users start experimenting with AI in democratized ecosystems, the ability to produce bespoke, responsive products will increasingly favour these grassroots innovators.

The global innovation ecosystem could be slowly transformed if Kerala and other hubs continue to export their democratized approach to AI and new technologies. The future of AI could be defined, invented and implemented by bottom-up networks of inventors and engineers, rather than large corporate platforms alone.

Several AI and policy experts have voiced the need to build a social license for AI, including new incentive structures to encourage state and private actors to align the development and deployment of AI technologies with the public interest.

In this context, the World Bank and international foundations, should discuss how to empower and oversee democratized innovation ecosystems, the "grassroots," in their effort to design and deploy AI for solving local social problems. By empowering "AI from the streets," our cities could become globally connected, yet locally inventive and inspired by a diversity of knowledge and vision about our shared digital futures. Echoing Gibson, the future is already here. But we need to distribute its promises more evenly.

Collective empowerment through bottom-up, inclusive learning and social experimentation: Designing mechanisms ("feedback loops") for sustained interactions between decision-makers, experts, and citizens, starting at the upstream end of research and development, could yield significant dividends in exposing and improving the distributive implications of innovation. Such model conceives policy as collective experimentation. But the experimentation is now at the technological level as well. Situations emerge or are created which allow to try out socially-responsible technological innovations and to learn from them, i.e. experimentation. Society becomes a laboratory, one could say, as we witness in the experience of the Indian State of Kerala that built resilience to ecological and health (Covid-19) crises by betting on inclusiveness with bottom-up innovation ecosystems. Here, however, the experimentation does not derive from promoting a particular technological promise, but from goals constructed around matters of concerns and that may be achieved at the collective level. Such goals will often be further articulated in the course of the experimentation. The regime of collective experimentation depends on investment of effort from a diversity of actors who are willing to engage in innovation processes because they are concerned about a specific issue.

In the human capital context, we should experiment with "AI from the street," giving diverse communities within SAR countries/cities an opportunity to learn how to turn their data, ideas and designs into AI innovation. What if cities in the South Asia region could be globally connected, yet locally inventive and inspired by a diversity of knowledge and vision? Connecting the success stories, learning and experience of diverse decentralized innovation ecosystems – from Kerala to Kathmandu to Dhaka to Shenzhen – may help build on positive collective experimentation. Along the way, new forms of collaboration, skill-deployment, regulatory and policy approaches could be tested. The value of a tech-enabled civic culture which relies on bottom-up information sharing, public-private partnerships, 'hacktivism' and participatory collective action, could attract interest from key stakeholders seeking to emulate these approaches. Diaspora communities may provide critical know how, mentoring, funding, and networks to local innovators. The opportunity would be a renewed interest by entrepreneurs and governments in 'steering innovation' to tackle societal challenges.

Finally, the World Bank needs to re-invigorate a discussion about how to build a social contract for AI and converging technologies for human capital, including new incentive structures to encourage state and private actors to align the development and deployment of AI technologies with the most vulnerable populations' interest. Technologies of humility,[173] a term coined by Sheila Jasanoff from Harvard Kennedy School, will depend on more than safe algorithms and well-curated data. It will rest on the humility of those who thought they could fully master AI, and the empowerment of others who can imagine locally beneficial intelligent designs.

Though the goal is ambitious, it is the only way to shape technological convergence so that it empowers vulnerable populations, protects human rights, and meets the ethical needs of a digitalizing and globalizing world.

[173] https://sciencepolicy.colorado.edu/students/envs_5100/jasanoff2003.pdf

SUMMARY OF INSIGHTS & RECOMMENDATIONS

A PARADIGM SHIFT FOR HUMAN CAPITAL

□ We enter a phase of technological convergence where AI can automate other dual-use technologies and industrial platforms that are critical to human capital and civilian populations' survival.

□ Such converging technologies lead to a new paradigm where our daily lives – our biometrics, emotional, behavioural & biological data – are free material for commercial and state surveillance. This form of automated profiling is already used in South Asia (in employment, education, health-surveillance), creating new power asymmetries, accountability challenges and needs for oversight.

□ AI is a transformative paradigm for human capital because it is essentially replacing the existing epistemic methods developed by humans and societies to produce knowledge and assess its value. Knowledge-production is increasingly automated by algorithms away from our explanatory scrutiny.

□ The convergence of AI and powerful data-capture technologies is giving rise to "affective computing," algorithms that can analyse us, predict our behaviours and emotions with drastic impact for education and employment. Youth and adults will have to adapt to an era, not only of automation, but also of competition with algorithms for cognitive and creative performance. Population subgroups could be excluded from economic flourishing, owing to both, lack of jobs and relevant skills.

CURRENT APPLICATIONS IN HUMAN CAPITAL – "QUICK WINS," OPPORTUNITIES AND RISKS

□ Converging technologies are drastically transforming human capital through the optimization of digital platforms for health (remote diagnosis), education (personalized learning) and social protection (social targeting). Under robust oversight, audit and human rights impact assessment, such optimization may provide quick wins for human capital in the South Asia region.

□ In human capital, AI and converging technologies can automate behavioural monitoring in education (emotional AI-based learning) & health (biometrics capture for stunting diagnosis). This prospect may improve remote access to human capital services, but comes with harmful data and algorithmic biases. Such biases and other failures in predictive analysis may lead to new patterns of exclusion and discrimination (intended/malicious or unintended).

□ Automated behavioural and emotional analysis in education and health may lead to new forms of data-commodification and exploitations, seriously undermining human rights, as well as amplifying cybersecurity threats to civilian populations. Such forms of citizens' profiling and micro-targeting is already used by digital platforms, data-brokers and intelligence corporations for social scoring. There is rising potential to amplify powerful existing divides, harm digital underclasses, with corrosive impacts for young women, minorities, climate and conflict-migrants and other vulnerable populations.

□ The use of AI-led behavioural and emotional data-analysis in education has corrosive implications for child rights. Legal reflections should centre on issues of proportionality, data-purpose and data-minimization. Weak scientific foundations and predictive value of AI emotional analysis in learning should also be addressed. The issues of human flourishing, social development and child educational benefits should be recognized as they are in the UN Convention on the Rights of the Child. The Convention clearly states the need to act in the child's best interests (Art. 3), the child's right to freedom of thought (Art. 14), privacy (Art. 16), the right to develop full potential (§1 Art. 29), the child's right to liberty (§2 Art. 29), and the child's right to be protected from economic exploitation (Art. 32).

LONGER-TERM OPPORTUNITIES & RISKS FOR HUMAN CAPITAL

□ Further breakthrough will come from converging technologies in precision medicine, precision agriculture, AI-led knowledge production and smart urban infrastructures. Yet, two factors may lead to human capital erosion: 1) distributive inequalities in how such breakthroughs are benefiting the most vulnerable populations in SAR countries; 2) rising behavioural and biological/biometrics surveillance by states and large corporations. It is increasingly difficult to isolate converging technologies from their potential for surveillance by corporations, state and non-state actors. Regulatory frailty is a pervasive concern for human rights infringements.

□ The landscape of cybersecurity threats is also expanding as well as the civilian targets (populations' data) that form the attack surface. Global cybercrime targets include the manipulation of data-sets within civilian information infrastructure. Cyberattacks impact global health and food supply chains.

GOVERNANCE OF CONVERGING TECHNOLOGIES IS A PRIORITY FOR HUMAN CAPITAL

A Theory of No-Harm

□ Within the Bank, there are no agreed and stress-tested methods to assess the ethical, security and human rights implications of delegating some strategic elements of human capital to automated predictive technologies. Therefore, a substantial accountability gap exists when no methods and cross-sector collaborations have been conceived and deployed to anticipate unforeseen misuses and long-term impacts of AI and data-capture technologies on vulnerable populations.

□ There is an urgent need for a "Theory of No-Harm" to anticipate & weigh the human capital benefits of converging technologies against the costs to fundamental human rights. Normative foresight scenarios can help anticipate, prevent, mitigate and account for civilian and human rights harms. Human rights impact assessment can support World Bank teams in systematically monitoring for ethical breaches and violations of human rights on the ground, impacting vulnerable populations.

□ Safeguards should protect data-privacy and security, empowerment, agency and social cohesion. Legal reflections should centre on issues of data-proportionality, data-purpose and data-minimization. Audit should prevent algorithmic biases leading to patterns of discrimination/exclusion.

□ Developing mechanisms for mitigating critical data incidents would support a common "Theory of No-Harm" and help improve digital guidelines as we learn.

□ Governance frameworks urgently needs drastic improvements to avoid an accountability crisis in SAR countries. Regulation and oversight of converging technologies will become the next frontier, opposing secrecy and control against calls for multi-stakeholder engagement to discuss societal norms and a new social contract.

□ Precedents exist to reverse information and power asymmetries. The Save Silent Valley Movement India's powerful resistance movements contributed to force the Bank to adopt new environmental and social standards and suspend investment in large dams. Recent movements of resistance against the use of facial-recognition by law enforcement is a case in point.

□ When it comes to human capital projects, there is an urgent need to consider if there could be irreversible damage to empowerment and agency, especially amongst children, women and marginalized populations. Like in the context of environmental engineering, such potential for long-term harm is difficult to assess, but this is an even more critical reason for the Bank to define a "Theory of No Harm," along a theory of change in human capital.

Localized Innovation Capacity, Building Community Knowledge

□ Another vision exists where converging technologies empower the underserved through inclusive democratized innovation and social learning with robust community-based safeguards.

□ Democratized innovation ecosystems provide a collaborative education environment where students, teachers and citizens learn to use converging technologies (computation, engineering, biotech) to become entrepreneurs, agents of change, and innovators in their communities.

□ Human capital responses to crises, when they include converging technologies, have to be designed, deployed and socialized on a trust-based approach, with the assurance that such responses will not, in the present and future, further erode human capital through automation of employment or surveillance. Some regions and states – the State of Kerala in India is a good example – have promoted inclusive innovation shaped to local needs and ethical expectations. Kerala demonstrates at the core of its model that social cohesion and resilience will matter more than technological solutionism.

□ In the human capital context, we should experiment with "AI from the street," giving diverse communities within SAR countries/cities an opportunity to learn how to turn their data, ideas and designs into AI innovation. What if cities in the South Asia region could be globally connected, yet locally inventive and inspired by a diversity of knowledge and vision? Connecting the success stories, learning and experience of diverse decentralized innovation ecosystems – from Kerala to Kathmandu to Dhaka to Shenzhen – may help build on positive collective experimentation. Along the way, new forms of collaboration, skill-deployment, regulatory and policy approaches could be tested.