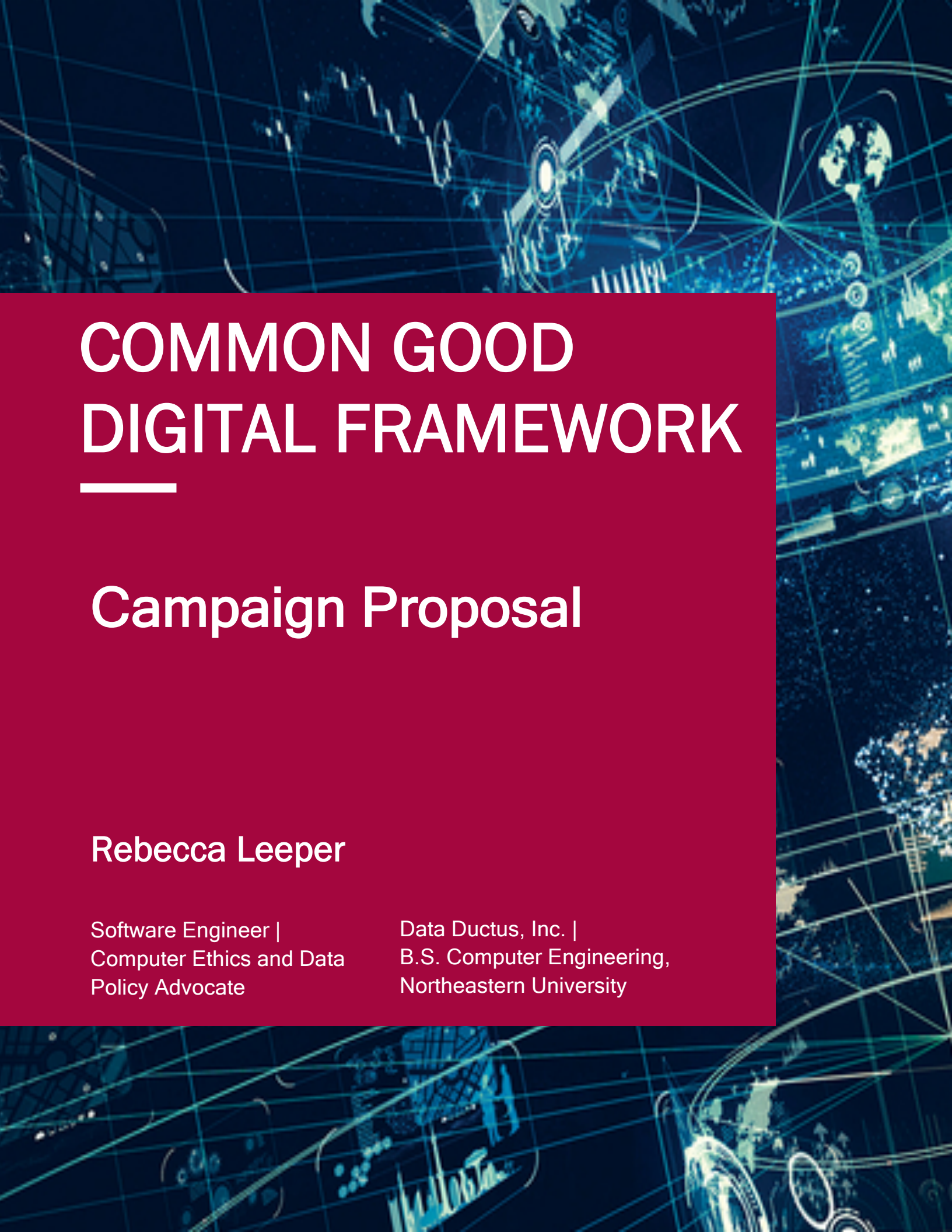

COMMON GOOD DIGITAL FRAMEWORK

Campaign Proposal

Rebecca Leeper

Software Engineer |
Computer Ethics and Data Policy Advocate

Data Ductus, Inc. |
B.S. Computer Engineering, Northeastern University

AI WORLD SOCIETY

Innovation Network

Common Good Digital Framework Action Plan

PURPOSE

The Common Good Digital Framework (CGDF) will serve as a platform to bring authoritative knowledge and raise awareness about violations of ethical values and standards by governments and large organizations.

The platform will monitor and alert against the misuse of Artificial Intelligence (AI), personal data, and neglect of cyber security. The objectives of the campaign are to stimulate and galvanize civil society towards the need to create new norms and regulations, and therein influence public and private AI and cyber policy.

FOCUS

The CGDF will primarily focus on AI bias, privacy (data security), and cyber security. It will specifically monitor and determine when digital activity and AI misuse violates common values and shared ethical principles and put forth a remediation action plan. In particular, the CGDF will identify areas where there are no regulations and common standards and will propose new courses of action. Secondarily, the CGDF will address and identify good cyber practices

that can serve as guiding models. Both findings will contribute to the generation of an Ethics and Practice Index designed for policy makers, Chief Technology Officers, and other digital actors. The Framework will examine both public institutions (governments) and private corporations. Some critical governments to observe include China, Estonia, some member states of the European Union, Israel, Japan, Russia, Saudi Arabia, and The United States. In the private sector, companies such as Alibaba, Amazon, Apple, Baidu, Facebook, Google, Microsoft, and Tencent will be prioritized.

MONITORING METHOD

The campaign will rely primarily on open source resources to stay up to date on digital activity – newspapers, magazines like Wired and Medium, journals and reports of authoritative institutions and news media. Secondarily, participation in key policy meetings, global summits, and international panels will both promote the Framework and provide insight into the digital activities of relevant actors. When possible, interviews will also be pursued, with the goal to create a network of trusted scholarly and practitioner advisors. In due time, the CGDF will form a board of expert advisors and relevant experts that will often be consulted and meet annually. Additionally, it will create a database of knowledge and learned people who can provide counsel and consultation.

RESPONSE METHOD

In response to identified violations, lack of regulatory frameworks, and the detrimental absence of commonly agreed norms and policies to the public well-being, the CGDF will generate a course of action and a policy recommendation by first reaching out to the relevant stakeholders. The policy recommendation will be circulated to all stakeholders and the CGDF network first for feedback and secondly to encourage corrective action by the offender. For example, should a country be violating its citizens rights, the policy presented may be for other states to apply sanctions or a trade embargo on that country. Following,

the CGDF will publicly publish the recommendation without disclosing the names of the violators. The purpose of this publication will be to draw public attention to the type of violations already occurring. In a later stage, if no corrective action was pursued, the CGDF will reveal the names of the perpetrators publicly, to be amplified by the CGDF network and its partners.

MEANS OF DISTRIBUTION

The campaign will utilize social media, a website, and journal publications.

PARTNERS

Partners will be expected to contribute data and evidence when governments and/or companies violate norms. In response to violations, partners will join together their voices, resources, and platform, serving as change agents and advocates.

Professor Denise Garcia of Northeastern University; Professor Woodrow Hartzog of Northeastern University; and CGDF Leader Juan Gallego. The CGDF will also seek to connect other young AIWS leaders and cyber advocates in the world.

<u>Potential Future Partners:</u>
Professor Stuart Russell, UC Berkley
President Joseph Aoun, Northeastern University
Dr. Eva-Marie Muller-Stuler, Chief Data Scientist and Leader Data Sceince Centre of Excellence, IBM
Sacha Alanoca, AI Policy Researcher at The Future Society
Maria Luciana Axente, PWC UK AI Commons
Aimee van Wynsberghe. Co-director Foundation for Responsible Robotics
Rebecca Distler, World Economic Forum Global Shaper
Yolanda Lannquist, Head of Research at The Future Society
Irakli Beridze, Head of the Centre for Artificial Intelligence and Robotics, UNICRI

James Butcher, Committee Member – Global AI & Ethics Initiative – IEEE
Representatives from other local and international NGOs

ACTION PLAN

1. Identify and assign an enthusiastic person for communications on social media for the CGDF
2. Identify and assign an enthusiastic person for outreach to connect other young people to join and contribute to the CGDF
3. Identify and seek out local and international NGOS to join the CGDF network
4. Identify and seek out relevant scholarly leaders to join the CGDF network
5. Identify and seek out leading cyber industry professionals to join the CGDF network
6. Identify and seek out relevant policymakers to join the CGDF network
7. Identify and seek out relevant conferences
8. Draft a report to present at AIWS Summit April 28-29, 2020 at Harvard University
9. Develop a rollout plan for publications and reports
10. Coordinate the campaign with the celebration of Plymouth 400th

Importance

The extensive use of the cloud, internet, and interconnected networks to manage and control infrastructure and the vulnerabilities associated with connectivity prone to hacking and interference pose significant security threats to human security (or the security and privacy of individuals), but also that of the state, and subsequently the international system, as this cyber domain can no longer be protected and observed by physical boundaries. The expansion extends to all aspects of a state's activity, including the private and public sectors rapidly adapting to the cloud and cloud computing environment for information storage, the internet being used to manage electric power generation, and the water levels in dams remotely by use of what is known as the Supervisory Control and Data Acquisition (SCADA) systems, and distribution networks for food, water, energy, transportation, healthcare, and financial services also depend heavily on the internet and information technology (IT). Moreover, the military depends on IT for command, control, and logistics, and modern-day weapons depend on GPS networks, and movements and actions of military forces are coordinated through networks that allow for immediate information exchange. This entire domain, which is interlinked with what is known as the critical infrastructure of a state, is highly vulnerable to a cyberattack.

Some Attacks on Critical Infrastructure

- BlackEnergy3 and CrashOveride : Ukraine Power Grid -2015 and 2016
- DragonFly/Energetic Bear: US Energy companies – 2011-2016
- Triton Attack : Saudi Petrochemical Plant – 2014-2017
- Agent.btz: US military networks – 2008

Some personal data leaks

- In 2016, 3 billion Yahoo accounts were hacked in one of the biggest breaches of all time.
- In 2016, Uber reported that hackers stole the information of over 57 million riders and drivers.
- In 2017, 412 million user accounts were stolen from Friendfinder's sites.
- In 2017, 147.9 million consumers were affected by the Equifax Breach.
- According to 2017 statistics, there are over 130 large-scale, targeted breaches in the U.S. per year, and that number is growing by 27 percent per year.

- 100,000 groups in at least 150 countries and more than 400,000 machines were infected by the Wannacry virus in 2017, at a total cost of around $4 billion.
- Attacks involving cryptojacking increased by 8,500 percent in 2017.
- In 2017, 5.4 billion attacks by the WannaCry virus were blocked.
- There are around 24,000 malicious mobile apps blocked every day.
- In 2017, the average number of breached records by country was 24,089. The nation with the most breaches annually was India with over 33k files; the US had 28.5k.
- In 2018, Under Armor reported that its “My Fitness Pal” was hacked, affecting 150 million users.
- Between January 1, 2005 and April 18, 2018 there have been 8,854 recorded breaches.

Other Cybercrime Statistics

- Total damage related to cybercrime is projected to hit 6 trillion annually by 2021
- There is a hacker attack every 39 seconds
- 60% of Americans have been exposed to fraud schemes
- 95 percent of breached records came from three industries in 2016: Government, retail, and technology

The Fourth Industrial Revolution (4IR) is rapidly emerging before our very eyes; science and technology are advancing at unprecedented rates that have culminated in what the World Economic Forum’s “Global Risks Report of 2017” calls the “twelve key emerging technologies” (WEF, 2017). Among the twelve, one technology in particular has the potential to transform warfare, impact healthcare, alter humanity and the environment. AI advancements can lead to the greatest benefits, but also poses the greatest risk. The most prominent risk is the advancement of AI for its implementation in autonomous weapons systems- machines that can arguably be used to take warfare to an unprecedented scale. While these weapons can be used to improve accuracy, efficiency, and in a capacity of self-defense, the technology, left unregulated, poses the biggest threat to security and the future of our world. The World Economic Forum (WEF) argues that future global governance over these AI-based weapons systems will depend upon “the rules, norms, standards, incentives, institutions, and other mechanisms that shape the development and

deployment of each particular technology". Cooperative, innovative efforts towards the development of progressive international norms and forward-thinking agreements are absolutely critical for the future of global governance of AI and autonomous weapons development.

The advancement and integration of AI also poses a significant threat in exacerbating existing biases. For example, AI used in facial recognition technology (FRT) has proven to be disproportionally inaccurate when identifying African Americans as compared to whites. In a study done by Joy Buolamiwini, a leading researcher challenging bias in decision making software, error rates, made up of mismatches and no registering of a face, for leading FR algorithms were significantly higher for people of color, particularly black females, than whites. In the study, the skin tones were classified as follows: people of color were classified as darker, whites were classified as lighter, women of color were classified as darker females, white women were classified as lighter females, etc. Microsoft's MSFT had an error rate of 12.9% for darker skin tones, 20.8% for darker females, and only a .7% error rate for lighter skin tones (. Face++ completed the trial with an error rate of16.5% for darker IDs, 34.5% for darker females, and 4.7% for lighter IDs. Lastly, IBM's software had error rates of 22.4% for darker skin tones, a significant 34.5%error rate for darker females, and for lighter skin tones a 3.2% error rate.

Some scholars attribute the algorithmic shortcomings and bias to the lack of diversity in the data training sets for the FRT and lack of representation of minorities in the stem field in general. Lastly, aggravating the situation, the federal and state mugshot databases used for FR have a higher number of minority -African American, Latino, and immigrant- headshots than white headshots. As one scholar notes, the FRT will be overused and underperform on people of color.

In sum, for these reasons, the Common Good Digital Framework will identify both good and poor digital practices to advocate for ethical guidelines, frameworks, and policy. This will create new norms, raising the sector ethos, and benefitting the common good. With this, security and trust will be increased for all of humanity.