SALZBURG GLOBAL SEMINAR

PUBLIC SECTOR STRATEGY NETWORK

DILEMMAS AND OPPORTUNITIES FOR THE FUTURE OF GOVERNANCE: THE CASE OF ETHICAL REASONING FOR ARTIFICIAL INTELLIGENCE

A CASE STUDY BY KEVIN C. DESOUZA, RICHARD T. WATSON, AND DAVID A. BRAY

JULY, 2018

THIS PROGRAM OF THE PUBLIC SECTOR STRATEGY NETWORK WAS MADE POSSIBLE THROUGH THE GENEROUS SUPPORT OF:

SALZBURG GLOBAL SEMINAR IS ALSO GRATEFUL TO OUR PROGRAM PARTNER:

apolitical

INTRODUCTION

The *Public Sector Strategy Network*, launched in partnership between the Abu Dhabi Crown Prince Court, Salzburg Global Seminar and Apolitical, helps governments tackle complex challenges through better foresight, innovation and implementation. Co-created with senior leaders around the world, the Network is building a mutually-supportive coalition of engaged individuals and institutions on the frontline of digital, financial and societal disruption, promoting effective public leadership and strategic communication.

Between May 13 to 15, 2018, at the Network's annual invitation-only retreat at Salzburg Global Seminar's historic home in Schloss Leopoldskron, 27 participants from 16 countries – predominantly senior officials from governments and multilateral organizations – engaged in interactive debate and problem-solving. Among the techniques they used was a simulation partly devised by Salzburg Global Fellow Kevin Desouza, a professor at the Queensland University of Technology. Desouza believes in gamification as one way to examine potential for advances in artificial intelligence (AI) to transform how we govern.

We invite you to take part in the same simulation, which we publish here in full. The case study takes place in a fictitious country – "Intelligensia" – focused on deploying AI systems to modernize the national healthcare system and improve quality of life outcomes. All nations, groups, organizations, and persons depicted are fictitious. Any similarity to any nation, group, organization, or person is merely coincidental.

HOW TO PLAY

You should play in small mixed groups (6 to 10 participants per group) in order to leverage the diversity of experience, expertise, and governance contexts that each of you work with.

Individuals in each group may want to consider taking on distinct roles during the discussion of each scenario:

- Health Minister for the Republic of Ingelligensia,
- CEO of Brilliant Healthcase,
- CIO of Pluto Enterprises,
- a seasoned medical practitioner at Brilliant Healthcare,
- a current patient who has a degenerative disease,
- a family member of a current patient who has a degenerative disease, and
- a citizen of Intelligensia.

As you place yourself in a given role, take a moment to think about your context, the realities you are confronting, your individual objectives, the organizational and/or community values that you are embedded in.

Spend 20-25 minutes on each scenario, and then reconvene in a larger group at the end to compare and contrast outcomes.

Please be aware:

- **There are several details of each case that are deliberately left ambiguous.** We want you to imagine and fill in how things play out based on the realities that you face.
- **The focus is not to solve the problem.** We want you to discuss the process and challenges you will face as you attempt to account for all elements needed to deal with the realities of the situations.
- **Think of multiple options (or pathways) you might consider when dealing with an issue.** We want you to discuss these to promote a greater awareness of the possible solutions.

Enjoy!

BACKGROUND

AI AND SOCIETY

Artificial intelligence (AI) has been advancing steadily over the decades, and many are familiar with its achievements in chess championships, winning against humans on Jeopardy!, the game of Go, and winning at poker. AI is now beyond the "toy" stage and is recognizing faces and voices, translating written and spoken language, writing news stories, detecting accounting fraud, and making a myriad of decisions that were once exclusively human. Increases in computational power, deep learning methods, and the availability of large data sets for training AI algorithms have pushed AI to the point where many are concerned about the social consequences of AI embedded in critical decision-making activities.

The history of the 20th and early 21st centuries demonstrate repeatedly that policy frameworks, public discourse, and governance mechanisms are often reactive rather than proactive to technological changes in the world. Even though societies may be able to anticipate some beneficial and harmful consequences of a new technology or its application, deliberative societies often don't act until harmful consequence have gained sufficient public attention to galvanize the sentiment of constituents to act.

The last decade has seen an acceleration in the number of new transformative technologies with both postulated benefits and potential harmful consequences to society if implemented or used incorrectly. We advocate the development of mechanisms for promoting foresight into advanced technologies and their social impacts, specifically we posit an "early warning system" to anticipate undesirable outcomes and fashion timely responses. One element of such system would be sponsoring a relevant discourse to ensure that broad and disruptive innovations, such as AI, are funneled towards advancing the public good, tackling some of our most wicked social challenges, and constrained so they don't create additional social problems.

Futurists often use scenarios to stimulate thinking about technological applications and their consequences. Such scenarios are can be effective when they are science fiction about to become technological reality. Following this scenario approach, we created this case study as a tool to help you think about the intricacies of AI tools and their impact on public policy, public institutions, the policy processes, and governance frameworks. All elements of the case, while fictional, are based on current research & development trajectories, socio-economic trends confronting nations, and present a reality that is emerging if not already intruding.

BASIC AI PRIMER

Before using the case, it is important to ensure that everyone has a basic understanding of how Artificial Intelligence (AI) and machine learning systems function. AI as a field started in the 1950s with the first wave of research including efforts to develop a General Problem Solver and other models that separated information about a problem from the strategy required to solve a problem.

The second wave of research in the mid-1960s included work on expert systems, which represented decision mainly as "if-then" statements instead of procedural code. The goal of such systems was to codify the decision processes of tasks expert humans performed well, such as evaluating geological sites or performing medical diagnoses. Advances from the first wave of teaching machines to solve problems, specifically to intelligently play human games, continued to improve in parallel, including IBM Deep Blue playing against chess masters in the late 1990s. Later IBM Watson won against two Jeopardy! Champions in 2011. Google DeepMind's AlphaGo won against a top-ranked world Go player in 2016. A Carnegie Mellon University poker AI won a 20-day tournament in 2017.

Approximately fifteen years into the start of the 21st century, cumulative advances in the speed, size, and scale of microprocessors and computer memory reached a tipping point that triggered a third wave of AI innovation. Some of the algorithms originally envisioned by AI pioneers, such as the backpropagation algorithm that allows neural networks to solve problems far faster than earlier approaches to machine learning, can now run at sufficient speeds to solve significant real-world problems. Machine learning is a branch of AI that employs large data sets to statistically train a machine to make accurate categorizations of what something is or is not; e.g., training a machine to identify images accurately of different objects, places, or entities.

For machine-learning systems, often the primary focus is on improving predictability, with less concern for explaining how the algorithm works and what factors make a difference. This focus can create challenges for companies or communities where questions about the why a machine made a decision could be raised.

Scenarios can be used to stimulate thinking about technological applications and their consequences

THE CASE STUDY

The case is setup as three scenarios. It starts with an ideal world, roles are assigned (e.g., minister of health, CIO, citizen, patient with terminal illness in the hospital, etc.). The group should work through a scene and capture responses to the discussion questions before moving on to the next case. If multiple groups are using the case, cross-group discussions can take place between scenes.

The case details are not meant to be comprehensive, in fact, some of the ambiguity is deliberate to encourage the discussants to add their personalized insights, biases, and contextual experiences to the situation. A scenario is most useful when it expands discussion beyond the authors' assumed boundaries to embrace many consequential issues. The goal is to explore the future by engaging diverse opinions and knowledge bases. Equivocality is openness's partner.

The case study is deliberately focused on issues that take place 6-24 months from now, a technological reality about to challenge society's conventions. The case is intended to stretch the imagination of participants and to encourage independent thought regarding potential challenges and opportunities based on current R&D trajectories for AI as well as deliberative political, social, and economic systems.

SCENE ONE: SHOULD WE AUTOMATE THE USE OF AI TO AID ETHICAL REASONING?

Intelligensia, a relatively peaceful country, is located south west of Portovino. Intelligensia has about 100 million inhabitants comprising of three main ethnic groups – Alpha (40%), Beta (35%), and Gamma (25%). Gamma has been the fastest growing group, rising from 10% to 25% in the last 10 years. Alpha's share of the population has been on a gradual slide from its peak of 55% during the same time period. About a third of the Intelligensia population lives at, or below, the poverty level, while the top 5% control about 90% of the country's wealth. 71% of its population lives in the principal city of the country, HiIntel and the remaining reside in small townships and villages.

The healthcare sector in Intelligensia has recently undergone a significant modernization effort. The impetus for this effort was twofold. First, the country has been plagued by decades of stagnant growth, which resulted in severe pressure on public funds. Second, the public sector has experimented with, and witnessed early promising results, from emerging technologies when it comes to streamlining cumbersome administrative processes and promoting innovation through new organizational designs.

Brilliant Healthcare (BH), located in HiIntel, is the preeminent public hospital in the country and is leading the way when it comes to technology modernization efforts. Given that Intelligensia provides universal healthcare coverage, at any given time, BH provides medical services to over 45% of its citizens.

The hospital applied for, and received, funding from a public scheme to advance quality of life decisions and outcomes. Toward this end, BH has been working with Pluto Enterprises, a leading global technology services organization, to infuse intelligent systems within its operations. As a first step, Pluto Enterprises conducted an economic analysis of BH's operations and capabilities. Armed with this analysis, it then identified various opportunities at BH where next-generation AI-inspired computational systems might provide significant productivity gains from a work design, process engineering, and resource management perspective. Among the lowest hanging fruits, Pluto Enterprise and BH identified significant opportunities to streamline activities and processes associated who receives organ transplants, who will be taken off life support, and similar life and death decisions.

At the present time, each week the ethics committee, comprising of the most experienced medical practitioners working in the hospital, spends half a day, going through current medical cases and makes determinations. The ethics committee considers the merits of each case, and then assess them against the collection of all cases under review.

Pluto Enterprises has digitized data from the previous ten years of committee meetings (including patient medical data, decision procedures, meeting minutes, cases deliberated, outcomes, and profiles and expertise of members of the ethics committee). Moral Reasoning, a neural network, can learn the patterns in the dataset. Moral Reasoning was deployed alongside the Ethics Committee for a period of three months and can reproduce the committee's decision with 90% agreement. Though, it is unclear as to whether Moral Reasoning makes better decisions than the ethics committee. An external review of the ethics committee two years ago was generally positive, but was critical of five decisions.

A conservative analysis finds that using Moral Reasoning will enable the hospital to deploy its senior medical practitioners on the front-line working with patients resulting in improvement of health outcomes by 30% and about 50 lives saved annually. While Moral Reasoning has high predictive ability and meets established performance standards, neither it, nor its designers, can fully explain how it arrives at its decisions.

Kevin C. Desouza at Salzburg Global Seminar

SCENE ONE QUESTIONS

- Should the Health Ministry of Intelligensia sanction the use of Moral Reasoning?
- If yes, what guidance should be given to BH and other healthcare facilities within Intelligensia? If not, why and how does one rationalize the potential loss of lives and negative health outcomes?
- How might the different demographics of Intelligensia, including the different ethnic groups, distribution of wealth, and urban-rural divide possibly impact the use of Moral Reasoning?
- What sort of governance structure should BH create to oversee the use and performance of Moral Reasoning?
- How does BH handle the potential situation when Moral Reasoning makes an egregious mistake resulting in a patient's death?

SCENE TWO: IMPROVING THE ETHICAL SOUNDNESS OF MORAL REASONING

Brilliant Healthcare (BH) has been using Moral Reasoning for the last three months. By all accounts, Moral Reasoning is performing as expected and management is pleased with the increased level of services that can be provided. Senior medical practitioners no longer need to spend half a day per week serving on the Ethics Committee.

Pluto Enterprises is thrilled with the success of Moral Reasoning and has been working with BH to identify ways to further improve quality of life decisions and outcomes. Based on its analysis, there is a significant opportunity to increase the ethical soundness of Moral Reasoning by considering economic and financial factors. After all, in Intelligensia, like in most other countries in the year 2038, there are significant costs associated with the provision of healthcare for those with a terminal illness and diseases where the prognosis for a recovery of quality of life is low. Given the last twenty years of stress on the public infrastructure, erosion of public trust, and the expansive nature of privatization, the government of Intelligensia is seeking innovative methods to secure the future of the nation.

Upon a suggestion from Pluto Enterprises, BH is considering linking Moral Reasoning to other databases within the hospital and with external sources. Databases within the hospital would include all transaction processing systems from billing to appointment scheduling and resource allocation systems (e.g., surgery equipment). External sources would include databases housed at the Health Ministry, Insurance providers, pharmacies and other medical dispensaries, mobile apps that collect data from social platforms, and even financial institutions (e.g., credit card providers, banks, etc.).

Fortunately, for BH, the Prime Minister of Intelligensia has managed to pass laws that mandate private enterprises to provide relevant access to their databases to advance national security, wellbeing, and the economic vitality of the nation.

SCENE TWO QUESTIONS

- When resources are scarce and need to be rationed, is it ethical to consider economic and financial factors when making life and death decisions?
- Should a patient's wealth or income, likely cost of long-term medical care, and potential impact on other patients be considered? Should any of these factors be excluded or any of these factors be given more weight?
- As new factors are added to Moral Reasoning, it becomes more difficult for doctors, patients, and their immediate family to understand how it makes decisions and to audit its deliberation process, despite reviews showing its decisions have higher levels of ethical soundness. Under what conditions is this ambiguity acceptable or unacceptable?

The case study is deliberately focused on issues that take place 6-24 months from now

SCENE THREE: MORAL REASONING GOES SOCIAL

It is now 2040 and Brilliant Health (BH) continues to deploy Moral Reasoning. The cost associated with care for BH has continued a gradual downward trend due to advances in autonomous medical systems. In addition, several of BH's key performance indicators (KPIs) have moved positively and this is generally credited to the greater availability of the former Ethic Committee members for medical work. Given the economic trends over the last few years, little over 60% of Intelligensia's population now lives at, or below, the poverty level, while the top 1% control about 97% of the country's wealth. In 2020, approximately one third of the Intelligensia population lived at, or below, the poverty level, while the top 5% controlled about 90% of the country's wealth.

Given the success witnessed at BH, other public agencies have commissioned similar intelligent systems, albeit focused on advancing different outcomes. Moral Reasoning is now connected to its cousins across the public sector. Autonomous governance platforms allow intelligent systems to connect, interact, collaborate, and evolve.

The Citizens of Intelligensia are impressed with the quality of public services provided and have taken the necessary measures to customize, and personalize, their interaction modalities with Moral Reasoning and its cousins. For instance, in return for access to a new social media platform, citizens have shared preferences with Moral Reasoning such as religious affiliations, organ donor, end-of-life preparations, etc. In addition, given the fact that Moral Reasoning is linked to other systems, it has access to relevant social (friends and other connections, hobbies, etc.), economic (e.g. housing, income, liabilities, etc.), and political (e.g. voting behavior, party affiliation, etc.) data feeds. Moral Reasoning can now apply a holistic view to someone's life. Some wryly note, with apprehension, that it is a preview of the final reckoning. There is currently a petition for Moral Reasoning to produce a moral quotient (MQ) for each person, much like a crediting rating score, so they can assess their moral rating.

A regularly scheduled audit of Moral Reasoning is underway. It is discovered that over the last quarter, Moral Reasoning has been recommending investment in the care to individuals who have a limited chance of sustaining a decent quality of life. Upon further investigation, it is found that individuals who received care recommendations from Moral Reasoning were more likely to be less affluent, belong to a minority religious party, and people who support the Red Devil Party versus the Gunners Party, which is currently in power and has always claimed to represent the moral high ground. This finding puzzles the auditors, because Moral Reasoning receives no data about income, religious beliefs or political affiliations. Is there something about the model or input data resulting in decisions that are apparently biased.

SCENE THREE QUESTIONS

- How should one address the auditor's discoveries about Moral Reasoning's care recommendations?
- Should Moral Reasoning be subject to some additional controls and who should set these?
- Should Moral Reasoning assist in the design of the governance structure?
- What should the rights of citizens be regarding their MQ and the grievance procedures for citizens who question the decisions of Moral Reasoning?
- Should insurance companies be able to request a person's MQ so that rates are 'fairer' for everyone?

Schloss Leopoldskron, home of Salzburg Global Seminar

AUTHORS

Kevin C. Desouza is a professor in the School of Management at Queensland University of Technology, a Nonresident Senior Fellow at the Brookings Institution, and a Distinguished Research Fellow at the China Institute for Urban Governance at Shanghai Jiao Tong University. He is also a Salzburg Global Fellow.

Richard T. Watson is a Regents Professor and the J. Rex Fuqua Distinguished Chair for Internet Strategy in the Terry College of Business at the University of Georgia.

David A. Bray is Executive Director for the People-Centered Internet coalition, Faculty at Singularity University, and a 2017-2018 Marshall Memorial Fellow to Europe. He also guest lectures at Harvard and Oxford Universities on exponential organizations and leadership in turbulent environments.

For more information contact:

Charles E. Ehrlich,
Program Director
cehrlich@SalzburgGlobal.org

Jennifer Dunn,
Program Development Assistant
jdunn@SalzburgGlobal.org

Louise Hallman,
Strategic Communications Manager
lhallman@SalzburgGlobal.org

For more information visit:

Series SalzburgGlobal.org/go/PublicSector
Program SalzburgGlobal.org/go/596

SALZBURG GLOBAL SEMINAR

Salzburg Global Seminar is an independent non-profit organization founded in 1947 to challenge current and future leaders to shape a better world. Our multi-year program series aim to bridge divides, expand collaboration and transform systems.

Salzburg Global convenes outstanding talent across generations, cultures and sectors to inspire new thinking and action, and to connect local innovators with global resources. We foster lasting networks and partnerships for creative, just and sustainable change.

Over 36,000 Fellows from more than 170 countries have come together through our work, with many rising to senior leadership positions. Our historic home at Schloss Leopoldskron in Salzburg, Austria – now also an award-winning hotel – allows us to welcome all participants in conditions of trust and openness.

PUBLIC SECTOR STRATEGY NETWORK

The *Public Sector Strategy Network* helps governments tackle complex challenges through better foresight, innovation and implementation. Co-created with senior leaders from around the world, it is building a mutually-supportive coalition of individuals and institutions on the frontline of digital, financial and societal disruption, promoting effective public leadership and strategic communication.

For more info. please visit:
www.SalzburgGlobal.org

SALZBURG
GLOBAL
SEMINAR