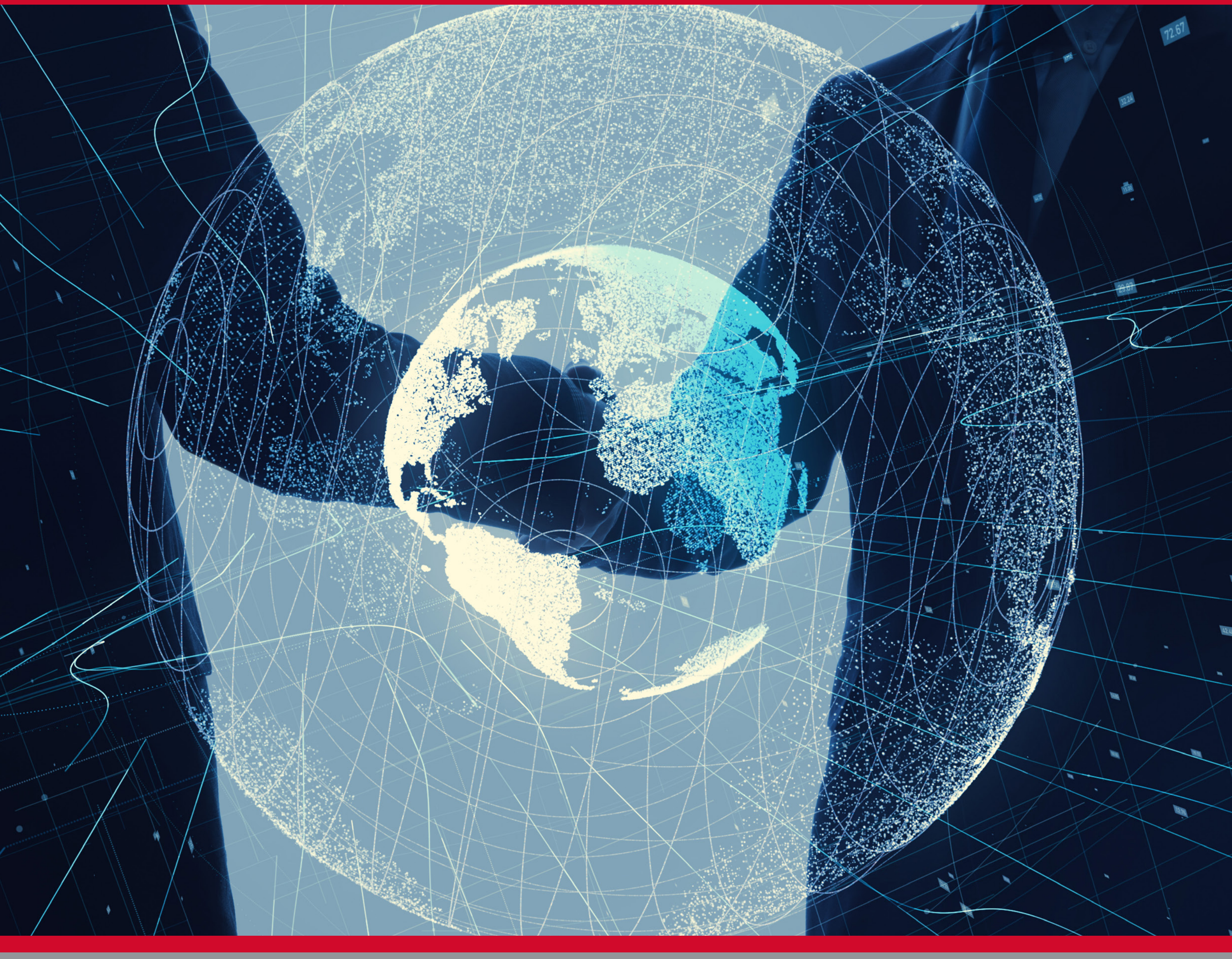
Diplomacy in the Age of Artificial Intelligence

Dr. Corneliu Bjola
Head of the Oxford Digital Diplomacy Research Group

أكاديمية الإمارات الدبلوماسية
EMIRATES DIPLOMATIC ACADEMY

EDA WORKING PAPER

January 2020



The Emirates Diplomatic Academy (EDA) is an autonomous federal entity of the United Arab Emirates (UAE) delivering high-impact diplomatic training, and disseminating research and thought leadership that furthers the understanding of diplomacy and international relations.

Emirates Diplomatic Academy
Hamdan Bin Mohammed Street
Al Hosn Palace District
P.O. Box 35567
Abu Dhabi, United Arab Emirates

http://eda.ac.ae/

research@eda.ac.ae

Dr. Corneliu Bjola

Head of the Oxford Digital Diplomacy Research Group

Dr. Corneliu Bjola is Associate Professor in Diplomatic Studies at the *University of Oxford* and Head of the Oxford Digital Diplomacy Research Group. He also serves as a Faculty Fellow at the *Center on Public Diplomacy* at the *University of Southern California* and as a Professorial Lecturer at the *Diplomatic Academy of Vienna*. His research focuses on issues related to the impact of digital technology on the conduct of diplomacy with a focus on strategic communication, digital influence and methods for countering digital propaganda.

He has authored or edited six books, including the twin volumes on *Countering Online Propaganda and Violent Extremism: The Dark Side of Digital Diplomacy* (2018) – listed in the Best New International Relations Books To Read In 2019 - and *Digital Diplomacy: Theory and Practice (2015).* He is currently working on a new project *"Diplomacy in the Age of Artificial Intelligence: Evolution or Revolution?"* examining the potential transformative impact of AI in consular affairs, crisis communication, public diplomacy, and international negotiations.

CONTENTS

Diplomacy in the Age of Artificial Intelligence

Executive Summary

- The global race to developing Artificial Intelligence (AI) capability is on with a clear focus on military, government surveillance, and economic applications. Studies have also started to investigate the AI potential for tackling some of the world's most challenging social problems and have found reasons for optimism that AI can improve conditions in a variety of social domains ranging from educational challenges to addressing issues of health and hunger. As with other technological revolutions, AI is bound to have far-reaching consequences in every corner of our societies, and diplomacy, by necessity, cannot escape its gravitational pull.
- The issue regarding the transformative impact of AI on diplomacy bears three interconnected questions: first, what exactly AI means, how does it work, and what forms it takes? Second, to what extent is AI capable of revolutionising how policy-makers take decisions? Third, in what areas of diplomatic activity could AI make a difference, in what form and with what kind of risks? By connecting theories of AI development, decision-making, and institutional adaptation, the working paper develops an analytical framework for examining the transformative relationship between AI in diplomacy in a manner that is conducive to generating relevant policy insight concerning the potential for AI integration in the activity of MFAs and embassies and its likely impact on core diplomatic tasks and activities.

Key takeaways

- AI refers to the activity by which computers process large volumes of data using highly sophisticated algorithms to simulate human reasoning and/or behaviour.
- The "AI effect" helps explain moving expectations about technology: as AI brings a new technology into the common fold, people become accustomed to this technology, it stops being considered AI, and newer technology emerges.
- Diplomatic functions can still be traditionally pursued with existing means and resources with no AI support. The new potential that AI brings, however, is cognitive augmentation, improved effectiveness and speed.
- AI systems are sufficiently developed conceptually to provide decisional support for a broad spectrum of diplomatic tasks. They are likely to evolve to allow automation of routinised tasks and services at the operational and tactical levels, but they will likely be kept out of strategic decision-making for technical and ethical reasons.
- Structured decisions, descriptive analytics and procedural knowledge are the most likely entry points for AI adoption in the diplomatic field; they apply well to consular services and international negotiations.

- Degrees of sustainability of AI technological innovation, perceptions of AI liability for diplomatic engagement, ethical reflections on matters concerning human control and AI (geo)political implications as well as practical considerations related to surveillance constitute potential exit points that may lead Ministries of Foreign Affairs (MFAs) to limit, postpone or abandon AI adoption.

- A potentially revolutionary contribution of AI to diplomacy would be to assist the decision-maker to prescribe a course of action in a dynamic fashion; that is, by constantly and automatically adapting its recommendations based on continuous description, diagnostic, prediction, and action loops.

- Ministries of Foreign Affairs could deploy the TIID framework as a conceptual roadmap for designing, delivering and deploying AI solutions in diplomacy, that combines considerations about what the objective is (task improvement), how to accomplish it (innovation), with what resources (physical/digital integration) and in what institutional configuration (deployment).

1. What is Artificial Intelligence?

> "By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it." - Eliezer Yudkowsky (AI researcher)

1.1. Definitions

The term "artificial intelligence" was first coined by an American computer scientist, John McCarthy in 1956, who defined AI as "the science and engineering of making intelligent machines, especially intelligent computer programs".[1] In basic terms, AI refers to the activity by which computers process large volumes of data using highly sophisticated algorithms to simulate human reasoning and/or behaviour.[2]

Since 1950, the ability of a machine to act humanly has been evaluated based on the Turing test. To pass the test, a machine should be able to engage in a dialogue for at least five minutes in such a way that a human interrogator could not distinguish its behaviour from that of a human being. To act humanly, a machine would have to meet two conditions: (i) react appropriately to the variance in human dialogue and (ii) display a human-like personality and intentions. No machine has passed the test thus far, and some researchers believe that for mathematical reasons it would be actually impossible to program a machine which can master the complex and evolving pattern of variance which human dialogues contain.[3] Thinking humanly, on the other hand, would imply that the machine would be able to think like a person that is, it would be able to store, process, organise information so that it can solve problems and make inferences about new situations. Drawing on theories of cognitive theory, the field of cognitive computing has taken the lead in combining symbolic methods with statistical methods to explore how AI can reason and learn with vast amounts of data.[4]

Russell & Norvig use these two dimensions (reasoning and behaviour) to group AI definitions according to the emphasis they place on thinking vs acting humanly. The definitions on top in Fig 1 are concerned, for instance, with thought processes and reasoning, whereas the ones on the bottom address behaviour. The definitions in the left column measure success in terms of fidelity to human performance, whereas the ones on the right measure AI excellence against an ideal rational performance.[5]

Another approach to defining AI is by zooming in on the two constitutive components of the concept. Nils J.

1 John McCarthy, "What Is AI? / Basic Questions," 2011, accessed May 22, 2019, http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.

2 In simple terms, behaviour refers to the way in which people act to a particular situation in response to certain internal or external stimuli. The classical theory of human behaviour, the Belief-Desire-Intention model (BDI) (see Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge University Press, 1999). argues that individual behaviour is best explained by the way in which agents develop intentions (desires that the agent has committed to pursue) out of a broader range of desires (states of affairs they would like to bring about), which in turn are derived from a set of beliefs (information the agent has about the world). The way in which intentions are formed remain a matter of dispute between different schools of thought, with a traditional view emphasizing the role of rational reasoning (the rational dimension) , while others stressing the importance of internal mental processes (the cognitive dimension), or the social context in which this occurs (the social dimension).

3 J. Landgrebe and B. Smith, "There Is No General AI: Why Turing Machines Cannot Pass the Turing Test," June 9, 2019, http://arxiv.org/abs/1906.05833.

4 Kenneth D. Forbus, "AI and Cognitive Science: The Past and Next 30 Years," *Topics in Cognitive Science* 2, no. 3 (March 8, 2010): 345–56, https://doi.org/10.1111/j.1756-8765.2010.01083.x.

5 Stuart Russell and Peter Norvig, *Artificial Intelligence A Modern Approach Third Edition, Pearson,* 2010, 2, https://doi.org/10.1017/S0269888900007724.

Figure 1: *Definitions of Artificial Intelligence*

Human-like Behavior

Thinking Humanly "The exciting new effort to make computers think... *machines with minds,* in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning..." (Bellman, 1978)	**Thinking Rationally** "The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	**Acting Rationally** "Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998) "AI ... is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Rational Behavior

Nilsson defines, for instance, artificial intelligence as the "activity devoted to making machines intelligent" while "intelligence is that quality that enables an entity to function appropriately and with foresight in its environment".[6] Echoing Nilsson's view, the European Commission's High-Level Group on AI provides a more comprehensive understanding of the term:

> "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected

6 Nils J Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge: Cambridge University Press, 2010), 13, https://doi.org/10.1017/CBO9780511819346.

> structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal."[7]

In sum, while the concept of artificial intelligence continues to evolve, one could argue that the ambition to push forward the frontier of machine intelligence is the main anchor that holds the concept together. As the authors of the report on "Artificial Intelligence and Life in 2030" point out, we should not expect AI to "deliver" a life-changing product, but rather to continue to generate incremental improvements in its quest to achieve and possibly surpass human standards of reasoning and behaviour. In so doing, AI also sets in motion the so-called "AI effect": as AI brings a new technology into the common fold, people become accustomed to this technology, it stops being considered AI, and newer technology emerges.[8]

1.2. Approaches

Three distinct approaches inform the development pathways of AI. The first is *symbolic artificial intelligence,* also known as Good Old-Fashioned AI (GOFAI), which refers to the creation of expert systems and production rules to allow a machine to deduce behavioural pathways. IBM's Deep Blue, which defeated Garry Kasparov in chess in 1997, used such a symbolic approach.[9] DeepBlue mimicked the behaviour of a grandmaster by using an algorithm extracting value from a set of static rules designed by a human to interpret and provide the optimal response to a particular chess move. GOFAI's main advantage comes from the fact that the process by which the algorithm reaches a decision is transparent, verifiable and explainable. Knowing exactly what the rules are and how the algorithm puts them together in decision trees makes it possible to test and improve the effectiveness of the system in incremental steps. At the same time, therein also lies the main weakness of GOFAI: its inability to adapt without human intervention to new circumstances, given the fact that it must rigorously follow a memorised set of rules.

Connectionist or computational or non-symbolic approaches to artificial intelligence involve providing raw environmental data to the machine and leaving it to recognize patterns and create its own complex, high-dimensionality representations of the raw sensory data being provided to it.[10] In other words, connectionist systems are capable of learning from raw data without direct human intervention. This could happen via machine learning (ML), by which the machine can learn on its own using statistical models without being explicitly programmed, or via deep learning (DL), a more sophisticated form of ML that uses a layered structure

7 High-Level Expert Group on Artificial Intelligence, "A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines," 2019, 6, https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

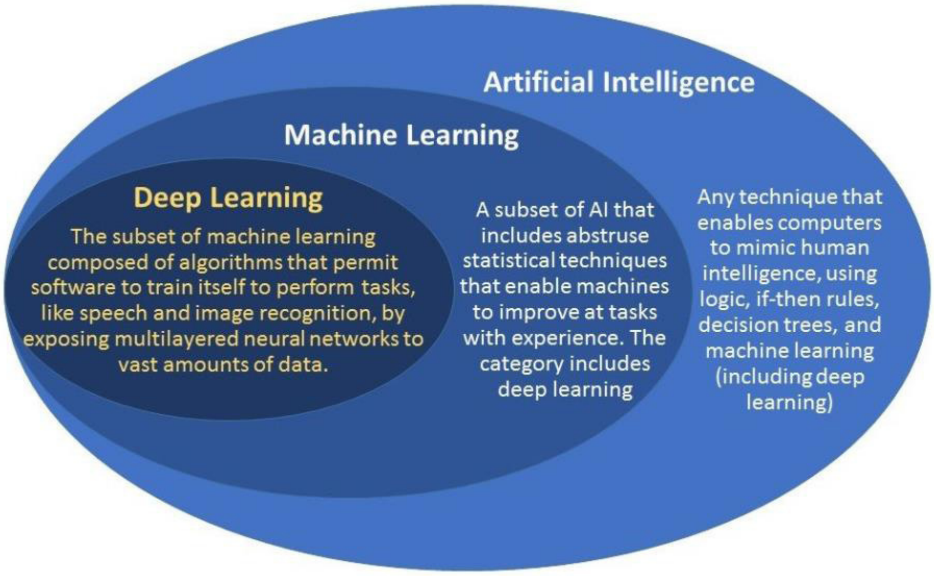
8 Peter et al. Stone, "Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, 52.," 2016, 12, https://ai100.stanford.edu.

9 Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review,* 2018.

10 Rhett D'souza, "Symbolic AI v/s Non-Symbolic AI, and Everything in Between?," OpenDeepTech, 2019, http://opendeeptech.com/symbolic-ai-v-s-non-symbolic-ai-and-everything-in-between.

of algorithms replicating an artificial neural network to process and classify information (see Fig 2). Facebook's news feed algorithm is an example of a machine learning approach seeking to promote "meaningful social interaction" by assigning greater weights to parameters that make a post personal, engaging and worthy of conversation.[11] Deep learning machines, on the other hand, go a step further and make possible, among other things, to automatically translate between languages or accurately recognize and identify people and objects in images.[12] AlphaZero, Google's AI program, is one of the most successful applications of DL. Starting from random play and given no domain knowledge except the game rules, AlphaZero convincingly defeated a world champion program in the games of chess and shogi (Japanese chess), as well as Go.[13]

Figure 2: *Machine vs Deep Learning*[14]



That being said, lack of interpretability of how algorithms reach decisions, poor generalisation of the results when using data outside the distribution the neural network has been trained on, and data inefficiency have exposed connectionist AI systems to legitimate criticism and have stimulated attempts to reconcile deep learning with symbolic artificial intelligence.[15] *Hybrid approaches* have thus emerged as a possible middle

11 Joshua Boyd, "The Facebook Algorithm Explained and How to Work It | Brandwatch," BrandWatch, 2019, https://www.brandwatch.com/blog/the-facebook-algorithm-explained/.

12 Bernard Marr, "10 Amazing Examples Of How Deep Learning AI Is Used In Practice?," Forbes, 2018, https://www.forbes.com/sites/bernardmarr/2018/08/20/10-amazing-examples-of-how-deep-learning-ai-is-used-in-practice/#4ee2c912f98a.

13 David Silver et al., "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play.," *Science (New York, N.Y.)* 362, no. 6419 (December 7, 2018): 1140–44, https://doi.org/10.1126/science.aar6404.

14 Meenal Dhande, "What Is the Difference between AI, Machine Learning, and Deep Learning," Geospatial World, 2017, https://www.geospatialworld.net/blogs/difference-between-ai-machine-learning-and-deep-learning/.

15 Marta Garnelo and Murray Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations," *Current Opinion in Behavioral Sciences* 29 (October 1, 2019): 17–23, https://doi.org/10.1016/J.COBEHA.2018.12.010.

ground solution, by combining minimal training data and no explicit programming with the ability to facilitate easy generalisation by deriving symbolic representation from supervised learning situations.[16] Hybrid systems define pathways for algorithmic computation by prioritising distributed vs localist representations of objects and concepts for solving low level and complex tasks.[17] Project Maven, founded by the US Defence Department, is a good example of how symbolic coding and deep learning have worked together to build effective AI capability to support drone missions in the fight against ISIS/Daesh.[18]

1.3. Types

In the same way that cars differ in terms of their quality and performance, AI programs also significantly vary along a broad spectrum ranging from rudimentary to super-intelligent forms. In consular and diplomatic affairs, the left side of this spectrum is already visible. At the lower end of the complexity scale, chat-bots now assist with visa applications,[19] legal aid for refugees,[20] and consular registrations.[21] More sophisticated algorithms are being developed by MFAs to either advance the spread of positive narratives or inhibit online disinformation and propaganda.[22] However, all these applications, regardless of their degree of technical sophistication, fall in the category of *'narrow' or 'weak' AI,* as they are programmed to perform a single task. They extract and process information from a specific dataset to provide guidance on legal matters and consular services. The 'narrow' designation for such AI applications comes from the fact that they cannot perform tasks outside the information confines delineated by their dataset.

By contrast, *general* AI refers to machines that exhibit human abilities ranging from problem-solving and creativity to taking decisions under conditions of uncertainty and thinking abstractly. They are thus able to perform intellectual activities like a human being, without any external help. Most importantly, *strong* AI would require some form of self-awareness or consciousness in order to be able to fully operate. If so, strong AI may reach a point in which it will be able not only to mimic the human brain but to surpass the cognitive performance of humans in all domains of interest. This is what Nick Bostrom calls *superintelligence,* an AI system that can do all that a human intellect can do, but faster ('speed superintelligence'), or that it can aggregate a large number of smaller intelligences ('collective superintelligence') or that it is at least as fast as a human mind but vastly qualitatively smarter ('quality superintelligence').[23]

That being said, strong AI, let alone superintelligence, remain merely theoretical constructs at this time, as all applications developed thus far, including those that have attracted media attention such as Amazon's

16 Chuang Gan, "The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision," in *Iclr 2019,* 2019.

17 For a dicussion of the differences between distributed and localist representation and their relevance for AI theory, see Asim Roy, "A Theory of the Brain - the Brain Uses Both Distributed and Localist (Symbolic) Representation," in *The 2011 International Joint Conference on Neural Networks* (IEEE, 2011), 215–21, https://doi.org/10.1109/IJCNN.2011.6033224.

18 Gregory C. Allen, "Project Maven Brings AI to the Fight against ISIS - Bulletin of the Atomic Scientists," The Bulletin of the Atomic Scientists, 2017, https://thebulletin.org/2017/12/project-maven-brings-ai-to-the-fight-against-isis/.

19 Visabot, "Immigration Attorney 2.0," https://visabot.co/.

20 Elena Cresci, "Chatbot That Overturned 160,000 Parking Fines Now Helping Refugees Claim Asylum," Guardian, https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay-refugees-claim-asylum-legal-aid.

21 Channel New Asia, "Most Singaporeans Do Not E-Register before Travelling," http://www.channelnewsasia.com/news/singapore/most-singaporeans-do-not-e-register-before-travelling-mfa-8775352.

22 Simon Cocking, "Using Algorithms to Achieve Digital Diplomacy.," Irish Tech News, http://irishtechnews.ie/using-algorithms-to-achieve-digital-diplomacy-a-conversation-with-elad-ratson-director-of-rd-at-ministry-of-foreign-affairs/.

23 Nick Bostrom, *Superintelligence : Paths, Dangers, Strategies,* First edit (Oxford: Oxford University Press, 2014), 63–69.

Alexa or Tesla's self-driving prototypes fall safely in the category of narrow AI. However, this may change in the near future, especially if quantum computing technology will make significant progress. Results from a large survey of machine learning researchers on their beliefs about progress in AI is relatively optimistic. Researchers predict AI will outperform humans in many activities in the next ten years, such as translating languages (by 2024), writing high-school essays (by 2026), driving a truck (by 2027), working in retail (by 2031), writing a bestselling book (by 2049), and even working as a surgeon (by 2053). Furthermore, they believe there is a 50% chance of AI outperforming humans in all tasks in 45 years and of automating all human jobs in 120 years.[24] With AI continuing to be dominated by 'narrow' applications for the foreseeable future,[25] memory-based functionality has emerged as an alternative benchmark for classifying AI.[26]

Case study #1: AI as an evolving Executive Assistant

Amb. Jane Doe is chairing the meeting of the UN General Assembly's committee on Disarmament and International Security. As the meeting goes on, a Type 1 (Reactive Machine) AI Assistant would be able to offer the Ambassador, upon request, relevant information regarding the issues on the agenda with accuracy and in real-time. A Type 2 (Limited Memory) machine may pick up cues from the conversation (e.g., topics that are discussed in more detail), form representations about the context of the conversation and use the cues to tailor its advice to the Ambassador insightfully and pro-actively. Drawing on memories of preceding meetings, a Type 3 (Theory of Mind) AI machine would be able to understand not only the conversation, but also the positions and strategies of the other participants to the meeting and advise the Ambassador accordingly. A Type 4 (Self-Aware) AI machine would likely be able to replace the Ambassador and chair the meeting itself.

- **Reactive Machines:** This form of intelligence is quite basic. It does not have the ability to form memories and cannot use past experiences to inform decision-making. It perceives the world directly and reacts to what is 'sees', but without having a representation of the world. The IBM chess program that beat Garry Kasparov in the 1990s is a good example of a reactive machine. It looked at the pieces on the chess board to assess the optimal move among various possibilities, but without any memory of its past moves and without any symbolic representation of the game of chess itself.
- **Limited Memory:** This category refers to AI systems that can use past experiences to inform current decisions, but these memories are transient. They cannot be converted into long-term experiences to be recalled and

24 Katja Grace et al., "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts," *Journal of Artificial Intelligence Research* 62 (July 31, 2018): 729–54, https://doi.org/10.1613/jair.1.11222.

25 White House, "Artificial Intelligence, Automation, and the Economy," 2016, https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED AI Economy Report.pdf.

26 Arend Hintze, "Understanding the Four Types of AI, from Reactive Robots to Self-Aware Being," The Conversation, 2016, https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616.

27 Vlad Savov, "Amazon Preparing a Wearable That 'Reads Human Emotions,' Says Report - The Verge," The Verge, 2019, https://www.theverge.com/circuitbreaker/2019/5/23/18636839/amazon-wearable-emotions-report.

used for taking decisions in similar situations. However, they have pre-programmed representations of the world to guide the application of short-term memories in decision-making. Self-driving cars use sensors, for instance, to form transient memories of instances of incoming traffic and road conditions. They then integrate these memories into pre-programmed representations of road transportation (road signs, traffic lights, driving rules) and take appropriate decisions about how to navigate safely.

- **Theory of Mind:** The next level of intelligence has a social dimension. It forms representations about the world by learning how to infer mental states about the entities inhabiting the world (their emotions, beliefs, intentions). This learning process facilitates the formation of long-term memories, which are used for behaviour adaptation in response to external stimuli. Amazon appears to be moving in this direction by working on a wearable device that will give Alexa, its flagship AI device, the ability to read human emotions and advise the wearer how to interact more effectively with others.[27]
- **Self-awareness:** The last stage of AI development takes place when the intelligence becomes capable not only of understanding the mental states of others, but also of itself. In other words, it develops consciousness. Instead of staying confined in the programmed setting that trains it to mimic human cognitive functions, self-aware AI learns how to think about itself and its surrounding environment. This form of intelligence approximates well Bostrom's three concepts of superintelligence mentioned above (speed, collective, quality). Echoing Bostrom's scepticism, it is probably safe to say that the development of self-aware AI remains unlikely for the time being without breakthroughs in AI hardware (computation speed, storage capacity, reliability) and software (duplicability, goal coordination memory sharing)[28] as well as in neuro-science (especially on the connection between brain circuitry and learning).[29]

28 Bostrom, *Superintelligence : Paths, Dangers, Strategies,* 71–74.

29 Shimon Ullman, "Using Neuroscience to Develop Artificial Intelligence.," *Science (New York, N.Y.)* 363, no. 6428 (February 15, 2019): 692–93, https://doi.org/10.1126/science.aau6595.

1.4. Key Points

- AI refers to the activity by which computers process large volumes of data using highly sophisticated algorithms to simulate human reasoning and/or behaviour.
- Three distinct approaches inform AI's development pathways: symbolic, connectionist and hybrid.The main difference between the three approaches lies with how the machine learns to react to its environment: by using pre-defined conceptual representations about the world, by developing such representations by itself or by using a combination of these two methods.
- Machine learning refers to the method by which the machine learns to understand the environment by using statistical models; deep learning relies on a layered structure of algorithms replicating an artificial neural network to process and classify information.
- 'Narrow' or 'weak' AI are applications programmed to perform a single task based on a specific dataset. General AI refers to machines that exhibit human abilities. Superintelligence refers to AI that can surpass the cognitive performance of humans in all domains of interest. All AI applications developed thus far fall into the 'narrow' category.
- Memory-based AI models vary by their capacity to form memories, to use past experiences to inform decision-making, and to infer mental states about others and themselves.
- "The AI effect": as AI brings a new technology into the common fold, people become accustomed to this technology, it stops being considered AI, and newer technology emerges.

2. AI and Decision-Making: Hype vs Reality

> "I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted." - Alan Turing

When reflecting on the implications of AI on our societies, it is perhaps useful to recall the famous statement of the former US Secretary of Defense, Donald Rumsfeld, concerning the role of epistemic framing on strategic thinking: "as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say, we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know [...] it is the latter category that tends to be the difficult ones".[30] This reasoning logic captures surprisingly well the current debate on AI technological developments. More specifically, there is substantial confusion around what AI can do conceptually, what it might be able to do on a future date, and around what we do not yet know whether AI might be able to do. To clarify these distinctions, this section takes a close look at the conceptual breakthroughs, promises and opportunities that AI has already or might generate in the area of decision-making. Knowledge is power *(scientia potestas est),* as Sir Francis Bacon once insightfully noted, but only if, one would be right to add, the power so gained leads to good decisions. AI has indeed the potential to revolutionize the way in which decisions are taken and in so doing, it has the power to generate important opportunities as well as challenges for policy-making in foreign affairs in general, and in diplomacy in particular.

2.1. Known Knowns

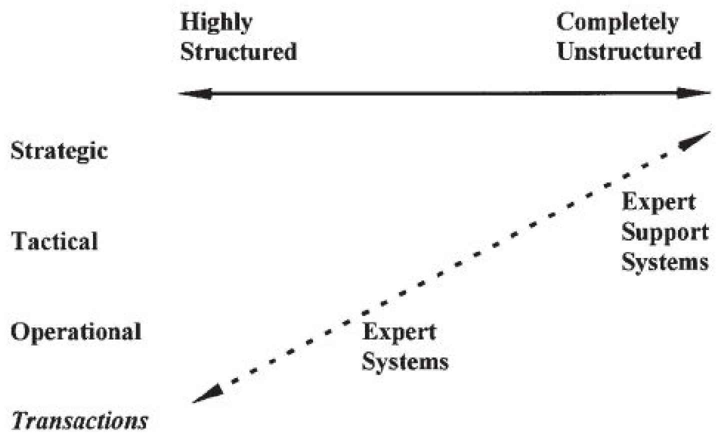
Conceptually speaking, what exactly can AI do and how efficiently can it assist decision-making? The concepts of expert system, decision structuredness, and embodied knowledge are central to addressing this dual question. Expert systems combine a knowledge base (a database of related information about a particular subject) and a set of rules to solving problems in a manner similar to a human expert. Knowledge-based systems can assist decision-making by serving in one of the following capacities:

- As an assistant invoked by the decision-maker to perform a specific task as part of a wider exercise.
- As a critic reviewing work already completed and commenting on aspects such as accuracy, consistency and completeness.

30 Donald H. Rumsfeld, "Defense.Gov Transcript: DoD News Briefing - Secretary Rumsfeld and Gen. Myers," U.S.Department of Defense, 2002, https://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636.

- As a second opinion executing the task so that its results may be compared with those of the decision-maker.
- As a consultant offering advice or opinions based on given information.[31]

Figure 3: *Use of expert systems for taking structured vs unstructured decisions*



The idea of harnessing the power of AI to facilitate decision-making in any of these roles is clearly appealing, but in order to work it must consider two possible constraints. The first one has to do with the degree of structuredness of the decision that is, how repetitive and routinised some of these decisions are (structured decisions), how novel and with no clear cut alternatives others may be (unstructured decisions), and how far along this spectrum a third category of decisions may fall (semi-structured decisions).[32] The assumption is that structured decisions are more amenable to AI modelling, while unstructured decisions are less prone to that, as they tend to rely on the judgement and insight of the decision-maker. Experimental results examining three organisational decision-making levels, i.e. strategic (unstructured), tactical (structured) and operational decisions (highly structured) provide empirical support to this view (see Fig 3).[33] More specifically, expert systems in a replacement role (i.e., taking decisions in place of a human) are effective at the operational and tactical decision levels but have limitations at the strategic level. Expert systems in a support role (i.e., giving advice or suggesting a solution to a problem) can help users make better decisions at all three decision making levels, but their effectiveness can only be fulfilled through their users.

31 John S. Edwards, "Expert Systems in Management and Administration - Are They Really Different from Decision Support Systems?," *European Journal of Operational Research* 61, no. 1–2 (August 25, 1992): 116, https://doi.org/10.1016/0377-2217(92)90273-C.

32 C. Burstein, F., & Holsapple, *Handbook on Decision Support Systems 1: Basic Themes* (Berlin: Springer-Verlag, 2008), 30.

33 J S Edwards, Y Duan, and P C Robins, "An Analysis of Expert Systems for Business Decision Making at Different Levels and in Different Roles," *European Journal of Information Systems* 9, no. 1 (March 19, 2000): 44, https://doi.org/10.1057/palgrave.ejis.3000344.

The second constraint involves the degree of interactivity between the decision-maker and the AI support system as the 'closed-world' picture of traditional expert systems understates the level of negotiation and engagement that exists and is expected to be cultivated between humans and machines.[34] As mentioned elsewhere,[35] by offering insights into the relationship between information technologies and social contexts,[36] social informatics (SI) research seeks to develop a better understanding of the ways in which people and information technologies interact and influence each other.[37]

Research into the type of knowledge that is embedded into AI systems reveals some interesting patterns in which the relationship between human and machine can develop (see Fig 4). "Expansion" involves, for instance, embodying teleological knowledge to extend cognition, thereby augmenting human work. "Emancipation" refers to embodying conditional knowledge to enact human cognition, thereby actuating human work (that is, reduces workload by completing work autonomously). "Equipping" is embodying procedural knowledge to scaffold human cognition, thereby assisting work (that is, reduce cognitive effort). "Expediting" involves embodying declarative knowledge to automate work (that is, reduce human workers' physical effort).[38]

To illustrate, an AI diplomatic system with declarative knowledge *(know-what)* would be able to collect, aggregate and store publicly available information about the positions of different parties to a negotiation. One with procedural knowledge *(know-how)* would be able to retrieve relevant information based on certain input criteria to assist negotiators in their work. An AI system with conditional knowledge *(know-when)* would be able to make sense of the conditions for using declarative and procedural knowledge such when the system may recognize the potential of a breakthrough in the negotiations and provide relevant information to make it happen. Finally, an AI system with teleological knowledge *(know-why)* would be able to understand the purpose, intention, or rationale of using declarative, procedural or conditional knowledge such when the negotiations serve to bring an end to a conflict, mitigate climate change or to improve conditions for financial governance.

In sum, from a conceptual perspective, AI systems are sufficiently developed to provide decisional support from a variety of positions (assistant, critic, second opinion, consultant). Their effectiveness varies, however, with the nature of the decision (semi- or structured decisions being more amenable to AI modelling) and the degree of interactivity between the decision-maker and the machine (embodied knowledge that sustains collaboration should be able to facilitate better engagement than knowledge that stimulates competition).

34 Yanqing Duan, John S. Edwards, and Yogesh K Dwivedi, "Artificial Intelligence for Decision Making in the Era of Big Data – Evolution, Challenges and Research Agenda," *International Journal of Information Management* 48 (October 1, 2019): 68, https://doi.org/10.1016/j.ijinfomgt.2019.01.021.

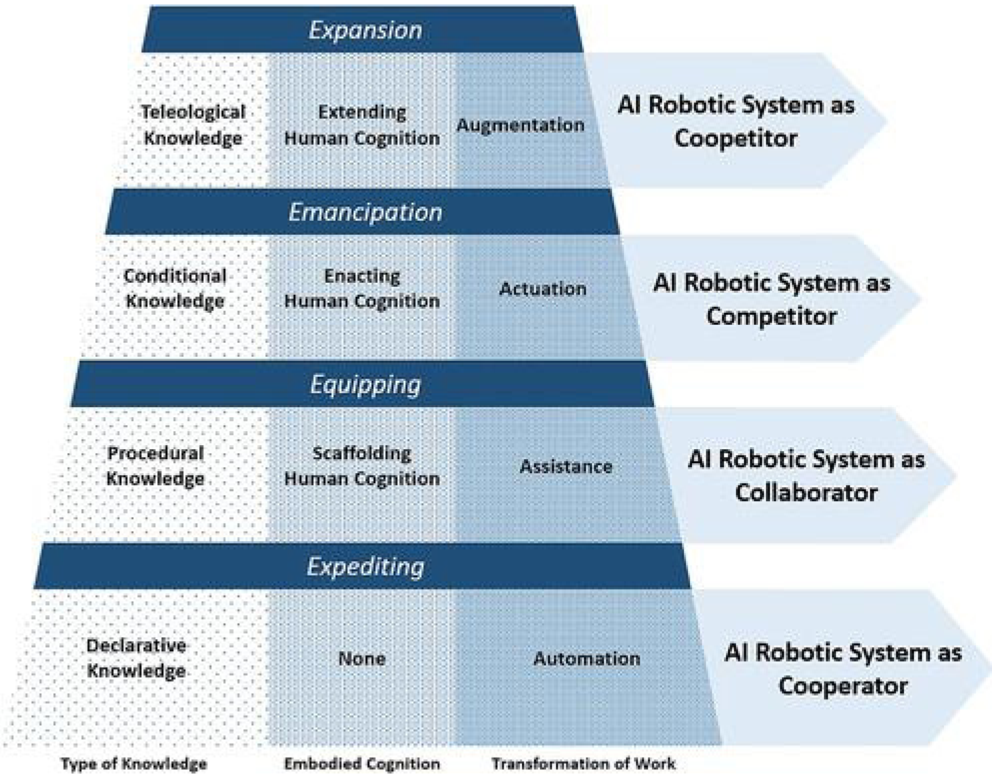
35 Corneliu Bjola, Jennifer Cassidy, and Ilan Manor, "Public Diplomacy in the Digital Age," *The Hague Journal of Diplomacy* 14, no. 1–2 (April 22, 2019): 84, https://doi.org/10.1163/1871191X-14011032.

36 Rob Kling, "What Is Social Informatics and Why Does It Matter?," *The Information Society* 23, no. 4 (2007); Steve Sawyer and Howard Rosenbaum, "Social Informatics in the Information Sciences: Current Activities and Emerging Directions," *Informing Science 3* 3, no. 2 (2000).

37 Ronald E. Day, "Kling and the Critical": Social Informatics and Critical Informatics," *Journal of the American Society for Information Science and Technology,* 58/4 (2007): 575; Mohammad Hossein Jarrahi and Sarah Beth Nelson, "Agency, Sociomateriality, and Configuration Work," *The Information Society* 34, no. 4 (2018).

38 L. G. Pee, Shan L. Pan, and Lili Cui, "Artificial Intelligence in Healthcare Robots: A Social Informatics Study of Knowledge Embodiment," *Journal of the Association for Information Science and Technology* 70, no. 4 (April 1, 2019): 365, https://doi.org/10.1002/asi.24145.

Figure 4: *Social informatics of knowledge embodiment*[39]



2.2. Known Unknowns

The idea that AI methods can be used to aid in the construction of theories of foreign policy behaviour has been around for some time.[40] Anderson & Thorson (1982) noted, for instance, that foreign policy is a goal-directed activity, i.e., "that foreign policy behaviour is not what a government 'does,' but rather it is what a government "does" described in the context of the ends toward which it was intended".[41] Therefore, the intentional aspects of foreign policy cannot be adequately explained by an exclusive reliance on causal factors. It also requires teleological explanations (see also Fig 4 above) to interpret the meanings of an

39 Pee, Pan, and Cui, 363.

40 Stuart A Bremer, "Computer Modeling in Global and International Relations: The State of the Art By My Definition, a Global Model Is a Multisector, Multi-Nation Computer Model of Long-Term Global Dynamics, While an Inter-National Relations (Hereafter IR) Model Is a Multination Model of in-Ternational Political Interactions," accessed May 30, 2019, https://journals.sagepub.com/doi/pdf/10.1177/089443938900700406.

41 Paul A Anderson and Stuart J Thorson, "Artificial Intelligence Based Simulations of Foreign Policy Decision Making'," *Behavioral Science* 27, no. 2 (1982): 178, https://onlinelibrary.wiley.com/doi/pdf/10.1002/bs.3830270211

action in light of the agent's particular situation and in light of the conventions, practices, and rules of the society.[42] Artificial intelligence, they further argued, could provide a potential vehicle for combining causal and teleological explanations of foreign policy behaviour. The interactive rule-based computer simulations discussed in their article, one regarding the Saudi foreign policy actions and the second examining President Kennedy's decision-making during the 1962 Cuban missile crisis, offered tentative empirical support to this view. Despite the ambition of these early attempts, little progress has been made thus far in introducing AI more systematically in the process of foreign policy decision making. However, this may change if new developments in AI technology are combined with a better understanding of the factors that shape decision-making in foreign policy.

The drive to harness the power of AI in foreign policy decision-making is informed by considerations about how to properly manage competing policy demands and institutional priorities. As Mintz and DeRouen point out, the decision-making environment places several constraints on the ability of foreign policy makers to compare, assess and pursue preferred courses of action.[43] Time is arguably a critical factor. Being able to react quickly, especially in the middle of a crisis, usually encourages decision-makers to intentionally condense the pool of alternatives, rely more on cognitive shortcuts, and search for a satisfactory rather than an optimal solution. The quality of the information is also critical for reaching sound decisions as without reliable information, it is difficult to compare alternatives, conduct a cost-benefit analysis, and assess utilities. Relatedly, ambiguity increases the complexity of the information environment, while familiarity downgrades it by favouring conclusions based on prior experience. Decisions are also interactive as they are being constantly shaped in reaction or anticipation of others' conduct and take place in a dynamic setting as options are often emerging and fading at a fast pace. The attitude toward risk and the capacity to cope with stress also influence information processing by causing decision makers to set unrealistic expectations or ignore certain information. Lastly, legitimate accountability can exercise a moderating effect on risk taking and thus improve the quality of the decisions. The nature of the organisational culture of the MFA (open vs closed, liberal vs rigid)[44] could also have significant influence on the way in which information is exchanged inside the organisation and by extension on how options are considered, evaluated and converted into decisions.

If we admit, like Pomerol, that any decision has its origin in a dissatisfaction arising from the difference between the current state of affairs and a more desirable, not yet existing one,[45] then the decision-making constraints identified by Mintz and DeRouen constitute possible liabilities which AI could address in order to

42 Anderson and Thorson, 177.

43 Alex Mintz and Karl DeRouen, *Understanding Foreign Policy: Decision Making, Understanding Foreign Policy: Decision Making* (Cambridge University Press, 2010), 25–30, https://doi.org/10.1017/CBO9780511757761.

44 Geert Hofstede et al., "Measuring Organizational Cultures: A Qualitative and Quantitative Study Across Twenty Cases," *Administrative Science Quarterly* 35, no. 2 (June 1990): 286, https://doi.org/10.2307/2393392.

45 Jean-Charles Pomerol, "Artificial Intelligence and Human Decision Making," *European Journal of Operational Research* 99, no. 1 (May 16, 1997): 4, https://doi.org/10.1016/S0377-2217(96)00378-5.

facilitate a meaningful resolution of the dissatisfaction gap. The analytical continuum developed by Gartner[46] (see Fig 5) offers a creative model for thinking about how to accomplish this task, which combines teleological and causal relationships as suggested above by Anderson & Thorson. The model proposes a four-dimensional typology of decision analysis, working together as a value escalator for unpacking the nature, the cause, the trend and the necessary course of action to be taken for addressing a particular problem. At the bottom of the value escalator lies descriptive analytics, which involves contextual mapping and the extraction of relevant information that can provide an accurate picture of the nature of the problem *(what happened?).* Diagnostic analytics allows the decision-maker to drill down or across the data in order to identify useful patterns and anomalies that can help isolate the root-cause of the problem *(why did it happen?).* Predictive analytics is all about forecasting likely courses of action and their possible implications by testing and validating certain assumptions about the nature and the cause of the problem *(what will happen?).* The last step, prescriptive analytics, encourages the decision maker to integrate the information gathered in the previous steps and use the result to determine the best course of action to be taken *(how to make it happen?).*

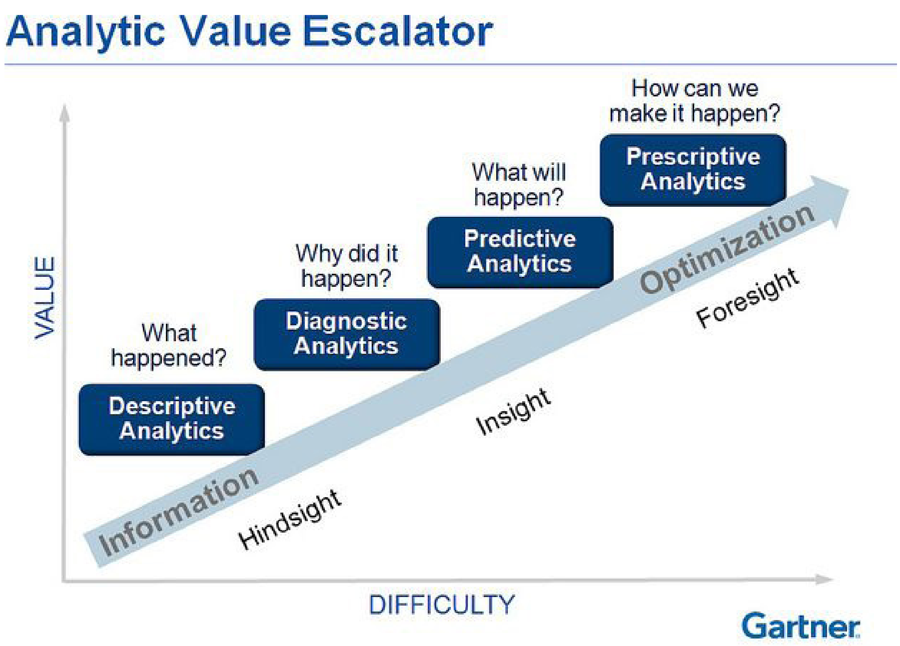
The higher we climb up on the value escalator, the better insight we gain for understanding the nature of the problem and for designing solutions to addressing it. The challenge for AI is to assist the decision-maker prescribe a course of action in a non-deterministic fashion that is, by constantly and automatically adapting its recommendations based on continuous description, diagnostic, prediction, and action loops. AI should be able to facilitate this process by enhancing the value of the following methods:

- *Data mining, data discovery and drill-down techniques* in the case of descriptive analytics. Being able to sift through large volumes of data from various sources, discover hidden relationships between variables and go deeper into the granular layers of data can have tremendous value for providing U.N. diplomats, for instance, with accurate personalised information (alerts, messages, visuals) when they prepare, conduct and/or chair a meeting.. AI can enhance these techniques by speeding up the process and improving the reliability and quality of data, two factors that Mintz and DeRouen found to be critical for decision-making.
- *Network, cluster and semantic analysis* in the case of diagnostic analytics. Understanding interdependencies between different events, identifying similar structures or clusters in the data, and extracting cognitive meaning from technical data can help map relevant causal relationships. AI could assist these methods by providing valuable insight into the formation and evolution of network patterns in data and interpreting their meaning using statistically-based machine learning algorithms as when diplomats may seek to assess

46 Gartner, "Magic Quadrant for BI Platforms. Analytics Value Escalator.," 2012.

the strength and validity of the signals they receive from their counterparts during a crisis. This could contribute to reducing the level of ambiguity in data, although not necessarily the degree of familiarity by which the information is processed.

Figure 5: *The Gartner Analytical Continuum*



- *Predictive modelling and multivariate* stats in the case of predictive analytics. This is the area in which machine learning is particularly suitable for identifying statistically relevant predictors that can be used for extrapolating trends and forecasting outcomes, a technique that could prove particularly useful to embassies for anticipating social or political tensions in their host countries.[47] In so doing, it may help mitigate risk concerns and provide a broader set of alternatives to be considered in anticipation to changing or evolving circumstances.
- *Simulation and complex event processing* in the case of prescriptive analytics. Monte Carlo simulation is one of the most commonly used techniques for performing risk analysis. It offers decision-makers a range

47 Abishur Prakash, "Algorithmic Foreign Policy," Scientific American Blog Network, 2019, https://blogs.scientificamerican.com/observations/algorithmic-foreign-policy/.

of probability ranked outcomes for multiple choices of action such as evaluating the long-term economic impact of Brexit by the U.K. government.[48] Complex event processing (CEP) is a method of tracking and analysing unfolding events in order to identify meaningful opportunities or challenges for a possible intervention and to respond to them as quickly as possible such as when information extracted from several stream events (an earthquake in country X, a diplomatic visit abroad of country X' head of state, and a history of recent political protests in the country) may alert the embassy in country X to expect a significant number of visa applications in short-term. Both methods are useful for deploying AI as a teleologically-oriented instrument of decision-making.

2.3. Unknown Unknowns

As the famous saying goes, "it is difficult to make predictions, particularly about the future". When thinking about the transformative impact of AI on our societies, this statement carries extra weight, as possible understatements could come at a heavy cost. When looking back, for instance, at the most important geopolitical feature of the nineteenth century, the Industrial Revolution, we can assess its implications with a fair degree of precision. As Drum points out, without the Industrial Revolution, there would be no rising middle class and no real pressure for democracy, no end to slavery due to the economy staying based on traditional subsistence agriculture, and no beginning of feminism. On the other hand, without the Industrial Revolution, there would be no colonization at scale because there would be no hard limit to a nonindustrial economy's appetite for raw materials, and no total war without cheap steel and precision manufacturing.[49] Performing the same evaluation for the AI (potential) revolution is, on the other hand, an arduous exercise partially because we lack good conceptual methods for benchmarking the scope and depth of the digital transformation and partially because of our direct involvement in this transformation as both observers and participants. Simply put, we do not know whether what we see happening around us as a result of the AI revolution represents the trunk of an elephant or the tail of a mouse, hence the difficulty of grappling with the "unknown unknowns" problem.

That being said, AI is here to stay so it is important to think carefully about its future evolution and possible implications. As Makridaksi suggests, four possible scenarios may follow in response to the question about how humans will adapt to the AI world.[50] *The Optimists* predict, for instance, that AI will likely have more positive implications, overall, than shortcomings, by ushering in a world of abundance, prosperity, well-being

48 H.M. Government, EU Exit: Long-term economic analysis, (Nov 2018), available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/760484/28_November_EU_Exit_-_Long-term_economic_analysis__1_.pdf

49 Kevin Drum, "Welcome to the Digital Revolution," *Foreign Affairs,* 2018, https://www.foreignaffairs.com/articles/world/2018-06-14/tech-world.

50 Spyros Makridakis, "The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms," *Futures* 90 (June 1, 2017): 50–52, https://doi.org/10.1016/J.FUTURES.2017.03.006.

and even potential immortality. *The Pessimists* reason that in a world dominated by 'strong' AI, the machines will take effective control of all important decisions with people becoming dependent on them and afraid of making their own choices. *The Pragmatists* fear the negative implications of AI, but they believe that with careful planning and regulation we should be able to learn to exploit the power of the machines to augment our own skills and stay a step ahead of AI, or at least not be at a disadvantage. Finally, *the Doubters* do not believe that AI is actually possible and that it will ever become a threat to humanity, largely because human expertise and especially creativity cannot be replicated and captured by formal rules.

Building on this line of thought, let us consider the heuristic value of these four scenarios in investigating the overall implications of AI development on the process of decision-making. The Optimistic scenario would likely expect AI to reach a point in which it could offer decision-makers high-quality assistance, covering the entire range of informational support from descriptive to prescriptive analytics, such when a diplomat would benefit from real-time information regarding the likely implications of different propositions during a negotiation. Such a development could be of a tremendous value for decision-makers as AI would help them make optimal decisions by reducing information gaps, improving the reaction time and keeping cognitive biases in check. At the same time, the quality of the AI support would much depend of the viability of its algorithms, especially if competing AI systems would seek to interfere with the machine's internal procedures for processing information (e.g., the party across the negotiation table would manipulate the evaluations the diplomat may receive from his/her AI assistant). If so, it would be important to understand how decision-makers will make their choices knowing that they could have only limited confidence in the integrity of the AI system **(Unknown Unknown #1)**. The Pessimistic scenario would anticipate growing resent to the involvement of 'strong' AI in the process of decision-making, as that could increase the risk of humans being left out of the loop as when decisions suggested the AI would be better trusted by the MFA than those offered by the diplomats at post. Reactions to this concern may take the form of regulation seeking to prohibit 'strong' AI from providing support for decision-making or to limit its use to delivering lower level of analytical input. However, one would expect that such ethical concerns might not be evenly shared by all actors with AI capacity, leaving those more committed to regulating AI at a disadvantage. If so, shall we then expect a regulatory race to the bottom for future AI development **(Unknown Unknown #2)**?

The Pragmatic scenario would predict a "controlled" integration of AI in the area of decision-making. AI development will be encouraged and supported via government policy or public-private partnerships but under conditions of close supervision so that any possible negative implications would be properly identified and addressed such as when all recommendations advanced by the AI during a diplomatic crisis would be vetted by a human. The strategy sounds sensible as it may help build public trust in the use of AI technology and limit inevitable challenges to policy decisions reached with AI support. One important issue under this scenario is the trade-off between AI acceleration and AI "explainability"[51] (i.e., understanding the internal logic by which AI generates outputs). Will this trade-off be sustainable and if not, will the pragmatic approach remain a credible option? **(Unknown Unknown #3)**? The Doubters scenario would probably be the least consequential for policy-making due to reservations concerning the possibility of AI to really influence the decision-making process, at least in its more sophisticated versions involving predictive and prescriptive analytics. One aspect worthy of consideration in this scenario is the cost of getting it wrong: what opportunity costs, if any, should be taken into account by those choosing to downplay AI impact **(Unknown Unknown #4)**? The feasibility of these scenarios much depends on the pace of technological innovation. AI breakthroughs in managing information in times of diplomatic crises or facilitating sound negotiation approaches will likely favour the Optimistic and perhaps the Pragmatic models, while slow progress in AI development such as error-prone or difficult to operate AI systems will rather support views aligned with the Doubters or Pessimistic scenarios.

51 Derek Doran, Sarah Schulz, and Tarek R Besold, "What Does Explainable AI Really Mean? A New Conceptualization of Perspectives," accessed June 3, 2019, http://amueller.github.

2.4. Key Points

- In order to make sense of AI's potential to revolutionize decision-making processes, we need to draw clear distinctions between what we know, what we do not know, and what we do not know we don't know.
- We know that AI systems are sufficiently developed conceptually to provide decisional support from multiple advice-giving positions, but their effectiveness depends on how structured the decisions are and whether they embody knowledge that sustains collaboration.
- We know that we don't yet know whether AI will be able to assist the decision-maker prescribe a course of action in a non-deterministic fashion that is, by automatically adapting its recommendations based on continuous description, diagnostic, prediction, and action loops. We might reduce this uncertainty by examining how AI may help improve techniques for conducting descriptive, diagnostic, predictive and prescriptive analytics.
- We do not know that we don't know whether what we see happening around us as a result of the AI revolution represents the trunk of an elephant or the tail of a mouse. Scenario-building (the optimists, the pessimists, the pragmatists and the doubters) can guide our thinking about how to identify some possible "unknown unknowns" (compromised AI integrity, regulatory race to the bottom, acceleration vs explainability trade off, opportunity costs of ignoring AI).

3. Diplomacy and AI

> "Some people call this artificial intelligence, but the reality is this technology will enhance us. So instead of artificial intelligence, I think we'll augment our intelligence." —Ginni Rometty (Chair, President, and CEO of IBM)

Riding the waves of growing interest about AI in IR and security studies,[52] the debate about the role of AI in diplomacy is also gaining momentum, although academic investigations are progressing rather slowly, without a clear analytical focus. Discussions about AI in the context of foreign policy and diplomacy often lack

52 Daniel W Drezner, "Technological Change and International Relations," *International Relations,* March 20, 2019, 004711781983462, https://doi.org/10.1177/0047117819834629.; Stoney Trent and Scott Lathrop, "A Primer on Artificial Intelligence for Military Leaders," *Small Wars Journal,* 2019, https://smallwarsjournal.com/jrnl/art/primer-artificial-intelligence-military-leaders.; Edward Geist and Andrew Lohn, *How Might Artificial Intelligence Affect the Risk of Nuclear War?* (RAND Corporation, 2018), https://doi.org/10.7249/PE296.; Greg Allen and Taniel Chan, "Artificial Intelligence and National Security," 2017, www.belfercenter.org.; Miles Brundage et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," February 20, 2018, http://arxiv.org/abs/1802.07228.

conceptual clarity, partly because of the persistent confusion between the two terms. While foreign policy is about the particular strategy a state decides to develop in pursuit of its national interests (who to make alliances with, what type of trade agreements to negotiate, or what policies to pursue in multilateral fora), diplomacy refers to the processes (representation, communication, negotiation) by which such a strategy is implemented.

To illustrate, the authors of a recent report suggest that a better understanding of the relationship between AI and diplomacy could come from building on the distinction between AI as a diplomatic topic, AI as a diplomatic tool, and AI as a factor that shapes the environment in which diplomacy is practised. As a topic for diplomacy, AI is relevant for a broader policy agenda ranging from economy, business, and security, all the way to democracy, human rights, and ethics. As a tool for diplomacy, AI looks at how it can support the functions of diplomacy and the day-to-day tasks of diplomats. As a factor that impacts the environment in which diplomacy is practised, AI could well turn out to be the defining technology of our time and as such it has the potential to reshape the foundation of the international order.[53] In fact, only the second dimension (AI as a diplomatic tool) falls into the realm of diplomatic activity. The other two dimensions (AI as a policy or external driver) have more relevance for foreign policy thinking.

Taking note of the fact that developments in AI are so dynamic and the implications so wide-ranging, another report prepared by a German think tank calls on Ministries of Foreign Affairs (MFAs) to immediately begin planning strategies that can respond effectively to the influence of AI in international affairs. Economic disruption, security & autonomous weapons, and democracy & ethics are the three areas they identify as priorities at the intersection of AI and foreign policy. The report thus draws a clear distinction between foreign policy objectives and diplomatic responses, but it fails to properly explain how the later could be developed. Although they believe that transformational changes to diplomatic institutions will eventually be needed to meet the challenges ahead, they favour, in the short term, an incremental approach to AI that builds on the successes (and learns from the failures) of "cyber-foreign policy", which, in many countries, has been already internalised in the culture of the relevant institutions, including of the MFAs.[54] By contrast, the authors of a report prepared for the Centre for a New American Security provide a more detailed perspective of how diplomacy and AI could assist national security objectives . For example, AI can help improve communication between governments and foreign publics by lowering language barriers between countries, enhance the security of diplomatic missions via image recognition and information sorting technologies, and support international humanitarian operations by monitoring elections, assisting in peacekeeping operations, and ensuring that financial aid disbursements are not misused through anomaly detection.[55]

53 DiploFoundation, "Mapping the Challenges and Opportunities of Artificial Intelligence for the Conduct of Diplomacy," 2019, https://www.diplomacy.edu/AI.

54 Ben Scott, Stefan Heumann, and Philippe Lorenz, "Artificial Intelligence and Foreign Policy," 2018, https://www.stiftung-nv.de/sites/default/files/ai_foreign_policy.pdf.

55 Michael C Horowitz et al., "Artificial Intelligence and International Security" (Washington D.C., 2018), https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security.

While offering a promising foundation for examining the evolving relationship between AI and diplomacy, current research nevertheless remains rather limited in its practical vision and analytical depth. There is a clear expectation for AI to demonstrate its potential as a tool in diplomacy, but it remains unclear what this potential means in practical terms and how it could be investigated theoretically. More to the point, it is important to map the areas of diplomatic activity in which AI could make a difference, to explore the nature of AI contributions, and to understand the risks they may entail for diplomatic work. Equally important, we should also seek to identify the main factors that can facilitate or hinder the adoption of AI by MFAs and to understand the conditions by which they may vary in scope and intensity. Building on the discussion on AI and decision-making, this section focuses on the diplomatic rather than the foreign policy realm and takes a close look at the opportunities and challenges of AI applications to consular services, crisis communication, public diplomacy, and international negotiations. The analysis will proceed in two steps, the first one examining the entry points for applying AI to diplomacy that is, the lead-in conditions that may induce MFAs to become receptive to the idea of adopting AI in their work. The second step will discuss the exit points to such an endeavour that is, the deterring conditions that may determine MFAs to limit, postpone or abandon AI adoption.

3.1. Entry Points

Drawing on the previous discussion examining the patterns by which AI can influence decision-making, this paper applies three conceptual benchmarks – the degree of decisional structuredness, the method of decision analysis and the type of embodied knowledge – for assessing the viability of the AI entry points in several key areas of diplomatic activity (see Fig 6). The assumption is that MFAs would more likely adopt a pro-active stance to integrating AI in their work if they can identify specific areas where AI can have the most effective influence on the task at hand, and then use these areas as testing grounds for further AI transformation. For a diplomatic activity mostly involving repetitive and routinised operations (consular services, pre-posting training) the first entry point would be, for instance, the task(s) that could be most effectively transformed through the introduction of an AI system that can assist with the taking of structured decisions, initially in a supportive role and later even as a replacement for the decision-maker.

The second entry point deals with the type of decision analysis that AI could best improve for the task at hand. A routinised diplomatic activity such as visa applications will hardly need predictive modelling supported by sophisticated algorithms to achieve its objectives, but it may benefit from data discovery techniques to assist with its descriptive analytics. At the same time, diplomatic projects with ambitious goals and limited budgets such as international negotiations would be better served by AI-driven simulations or complex event processing to navigate and manage policy uncertainties using prescriptive analytics. While the first entry point speaks to the issue of the practical relevance of AI in a particular diplomatic field, decision analysis zooms in on the practical contribution that AI can make to diplomatic decision-making processes.

Successful AI adoption also depends on how the knowledge so generated is converted into a form in which its value becomes evident for the organisation. The third entry point thus refers to the form of knowledge embodiment entailed by the AI adoption and whether it helps sustain collaboration vs stimulate competition between the users and the AI system. Routinised diplomatic tasks may encourage, for instance, forms of knowledge embodiment that may go as far as reducing workload by completing physical and cognitive work autonomously such in the case of passport renewals (emancipation). By contrast, other tasks for which human control is deemed more critical may favour forms of knowledge embodiment seeking to reduce only the cognitive effort (equipping) as in the case of international crises by improving the accuracy and reliability of the information received in real-time. Knowledge embodiment thus calls attention to the question of practical feasibility that is, of whether the organisation would be inclined to accept the AI system regardless of the value of its practical benefits.

The strength of the scientific and technological culture (S&T) in a society is also critical for understanding the potential success of AI development and institutional adoption. As "the expression of all the modes through which individuals and society appropriate science and technology"[56] SCT serves as an important structural determinant of the way in which technology in general and AI in particular could be seen as a natural assistant for knowledge expansion vs a potential competitor or even threat to societal development.[57] The 2019 Government AI Readiness Index pertinently acknowledges the significance of the S&T culture in fostering a conducive environment for AI adoption through the "Skills and Education" proxy indicator. As the report makes clear, good governance, reliable infrastructure and innovative public services are critical factors for AI development, but without a strong domestic S&T culture, governments will need to focus heavily on attracting and retaining foreign talent, which is already in short supply.[58]

As mentioned above, artificial intelligence is already making inroads in *consular affairs,* albeit at the lower end of the complexity scale via chat-bots and virtual assistants, which now provide support for visa

56 Benoit Godin and Yves Gingras, "What Is Scientific and Technological Culture and How Is It Measured? A Multidimensional Model," *Public Understand. Sci,* vol. 9, 2000, 44, https://archipel.uqam.ca/491/1/Sci_cult_Pub_Und_sc.PDF.

57 MW Bauer and A Suerdem, "Relating 'Science Culture' and Innovation," 2016, https://www.oecd.org/sti/097 - OECD Paper attitudes and innovation_v4.0_MB.pdf.

58 Oxford Insights, "Government Artificial Intelligence 2019 Readiness Index," 2019, 13, https://ai4d.ai/wp-content/uploads/2019/05/ai-gov-readiness-report_v08.pdf.

Figure 6: *AI entry points for diplomatic decision-making processes*

	Diplomatic Area			
Entry Point	Consular Services	Crisis Management	Public Diplomacy	International Negotiations
Decision Type				
Structured	xx		xx	x
Semi-Structured		x	x	xx
Unstructured		xx		
Decision Analysis				
Descriptive	xx	xx	x	xx
Diagnostic			xx	x
Predictive		x		
Prescriptive			x	x
Knowledge Type				
Declarative	x			xx
Procedural	xx	x	x	x
Conditional		xx	x	
Teleological				
AI Overall Potential[59]	Strong	Medium	Medium	Strong

59 The total score is not based on the number of ticks, but on how smoothly AI can be adopted in a particular area. Structured decisions, descriptive analytics and declarative knowledge are fairly suitable for AI adoption. Semi- or unstructured decisions, predictive analytics and conditional knowledge are more difficult to translate from an AI perspective, at least for the time being.

applications, consular registrations, and even legal aid for refugees.[60] Most of the work done by consulates covers procedural aspects related to the issuing of emergency passports, legalised documents as well as the provision of travel advice, guidance on governmental policy, or of information about trading and doing business in the home country. Occasionally, consulates are called upon to aid nationals in times of crises, which requires a different approach to making decisions.

Case study #2: AI as Digital Consul Assistant

The consulate of country X has been facing uneven demand for emergency passports, visa requests and business certifications in the past five years. The situation has led to a growing backlog, significant loss of public reputation and a tense relationship between the consulate and the MFA. An AI system trained with data from the past five years uses descriptive analytics to identify patterns in the applications and concludes that August, May and December are the most likely months to witness an increase of the demand in the three categories next year. AI predictions are confirmed for August and May but not for December. AI recalibrates its advice using updated data and the new predictions help consular officers manage requests more effectively. As the MFA confidence in the AI system grows, the digital assistant is then introduced to other consulates experiencing similar problems.

From a decisional perspective, consular services rely on highly-structured decisions, as they largely involve recurring and routinised operations based on clear and stable procedures, which do not need to be treated as new each time a decision has to be made (except for crisis situations, which are discussed further below). From a knowledge perspective, AI-assisted consular services may embody declarative (know-what) and procedural knowledge (know-how) to automate routinised operations and scaffold human cognition by reducing cognitive effort. This can be done by using data mining and data discovery techniques to organize the data and make it possible to identify patterns and relationships that would be difficult to observe otherwise (e.g., variation of demand for services by location, time, and audience profile). Consular services could be thus seen as a low-hanging fruit for AI integration as decisions are amenable to digitisation, the analytical contribution is reasonable relevant and the embodied knowledge favours collaboration between users and the machine.

60 Elena Cresci. "Chatbot That Overturned 160,000 Parking Fines Now Helping Refugees Claim Asylum." The Guardian, 2017. https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay-refugees-claim-asylum-legal-aid; Channel News Asia. "Most Singaporeans Do Not E-Register before Travelling." Channel News Asia, 2017. https://www.channelnewsasia.com/news/singapore/most-singaporeans-do-not-e-register-before-travelling-mfa-8775352 ; A shorter version of this argument is available at http://www.realinstitutoelcano.org/wps/portal/rielcano_en/contenido?WCM_GLOBAL_CONTEXT=/elcano/elcano_in/zonas_in/ari98-2019-bjola-diplomacy-in-the-age-of-artificial-intelligence

Digital platforms have emerged as indispensable tools for *managing diplomatic* crises in the digital age and for good reasons. They can help embassies and MFAs make sense of the nature and gravity of the events in real-time, streamline the decision-making process, manage the public's expectations, and facilitate crisis termination. At the same time, they need to be used with great care as factual inaccuracies, coordination gaps, mismatched disclosure level, and poor symbolic signalling could easily derail digital efforts of crisis management.[61] AI systems could provide great assistance to diplomats in times of crisis by helping them make sense of what it is happening (descriptive analytics) and identify possible trends (predictive analytics). The main challenge for AI is the semi-structured nature of the decisions to be taken. While many MFAs have

Case study #3: AI as Digital PD Assistant

The embassy of country X in London would like to conduct a public diplomacy campaign in support of one of the following policy priorities: increasing the level of educational exchanges of UK students in the home country, showcasing the strength of the military relationship between country X and the UK and boosting UK investments in the home country. As it has only £25,000 in the budget for the campaign, it needs to know which version can demonstrate better return on investment. Using social media data, an AI system will first seek to listen and determine the level of interest and reception (positive, negative, neutral) of the public in the three topics. The next step will be to use diagnostic analytics to explain the possible drivers of interest in each topic (message, format, influencers) and the likelihood of the public reacting to the embassy's campaign. The last step will be to run simulations to evaluate which campaign will be able to have the strongest impact given the way in which the public positions itself on each topic and the factors that may help increase or decrease public interest in them.

pre-designed plans to activate in case of a crisis, it is safe to assume that reality often defies the best crafted plans. Given the high level of uncertainty in which crisis decision-making operates and the inevitable scrutiny and demand of accountability to occur if something goes wrong, AI knowledge embodiment can work only if humans retain control over the process. As a recent SIPRI study pointed out, AI systems may fail spectacularly when confronted with tasks or environments that differ slightly to those they were trained for. Their algorithms are also opaque, which makes difficult for humans to explain how they work and whether they include bias that could lead to problematic—if not dangerous—behaviours.[62] This implies that a collaborative form of knowledge embodiment such as equipping would be best suited for this task as it can assist decision-makers

61 Corneliu Bjola, "How Should Governments Respond to Disasters in the Digital Age? | RUSI," The Royal United Services Institute (RUSI), 2017, https://rusi.org/commentary/how-should-governments-respond-disasters-digital-age.

62 Vincent Boulanin, "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk," SIPRI, 2019, https://www.sipri.org/publications/2019/other-publications/impact-artificial-intelligence-strategic-stability-and-nuclear-risk.

scaffold their human cognition so that they can better make sense and cope with the consequences of the crisis. The new data-based analysis system currently developed the German Federal Foreign Office promises to do exactly that: to use publicly available data on social, economic, and political developments to enhance human cognition on detecting likely crises ranging from pandemics to global migration flows.[63]

As data is turning into the "new oil", one would expect that the influence of digital technologies on *public diplomacy* to maximise interest in learning how to make oneself better heard, listened and followed by the relevant audiences. As the volume of data-driven interactions continue to grow at an exponential rate, one can make oneself heard by professionally learning how to separate 'signals' from the background 'noise' and by pro-actively adjusting her message to ensure maximal visibility in the online space, in real time. Making oneself listened would require, by extension, a better understanding of the cognitive frames and emotional undertones that enable audiences to meaningfully connect with a particular message. Making oneself followed would involve micro-level connections with the audience based on individual interests and preferences.[64]

At the operational level, new techniques such as visual enhancement, emotional framing, and algorithmic-driven engagement will be essential for digital diplomats to use in order to meaningfully communicate with their target audiences. AI could assist these efforts by providing reliable diagnostics of the scope conditions for impact via network, cluster and semantic analyses. Prescriptive analytics could also offer insight into the comparative value-added of alternative approaches to digital engagement (e.g., which method proves more impactful in terms of making oneself heard, listened and followed). On the downside, the knowledge so generated would likely stimulate a competitive relationship between the AI system and digital diplomats as most of the work done by the later could be gradually automated. However, such a development might be welcome by budget-strapped MFAs and embassies seeking to maintain their influence and make the best of their limited resources by harnessing the power of technological innovation.

Given the growing technical complexity and resource-intensive nature of *international negotiations* it is hardly surprisingly that AI has already started to disrupt this field. The Cognitive Trade Advisor (CTA) developed by IBM aims to assist trade negotiators dealing with rules of origin (criteria used to identify the origin /nationality of a product) by answering queries related to existing trade agreements, custom duties corresponding to different categories of rules of origin, and even to the negotiating profiles of the party of interest.[65] CTA uses descriptive analytics to provide timely and reliable insight into technically complex

63 Marcel Dickow and Daniel Jacob, "The Global Debate on the Future of Artificial Intelligence. The Need for International Regulation and Opportunities for German Foreign Policy," 2018, 6–7, https://www.swp-berlin.org/fileadmin/contents/products/comments/2018C23_dkw_job.pdf.

64 Bjola, Cassidy, and Manor, "Public Diplomacy in the Digital Age," 87.

65 Maximiliano Ribeiro Aquino Santos, "Cognitive Trading Using Watson," IBM blog, 2018, https://www.ibm.com/cloud/blog/cognitive-trading-using-watson.

issues that would otherwise require days or possibly weeks for an experienced team to sort out. It does not replace the negotiator in making decisions, nor does it conduct negotiations by itself, or at least not yet. It simply assists the negotiator in figuring out the best negotiating strategy by reducing critical information gaps. The competitive advantage that such a system could offer negotiators cannot be ignored, although caveats remain for cases in which negotiations would involve semi-structured decisions such as climate negotiations or the Digital Geneva Convention to protect cyberspace. The problem for such cases lies with the lower degree of data veracity (confidence in the data) when dealing with matters that can easily become subject to interpretation and contestation, hence the need for stronger human expertise and judgement to assess the value of competing courses of action in line with the definition of national interests as agreed upon by foreign policy makers.

In sum, the three entry points provide a useful analytical map for understanding the relevance, feasibility and contribution of AI to diplomatic work. As an expression of the societal support or reservation for technology, the Science and Technology Culture of a society could amplify or hinder the intensity with each these factors may influence AI adoption by MFAs. As the cases examined above make clear, structured decisions, descriptive analytics and equipping embodied knowledge present themselves as the most likely entry points for AI adoption. They apply particular well to consular services and international negotiations as most of the decisions in this area rely on established procedures and require a basic form of interaction between users and the AI system. Crisis management and public diplomacy are slightly more challenging domains due to the dynamic nature of their environment and /or the higher expectations they set for the human-machine relationship. The four categories of diplomatic activity examined above cover areas of general interest for MFAs, but the 'entry point' model could be easily generalised to other forms of diplomatic engagement as well.

3.2. Exit Points

In the same way in which certain factors may facilitate AI adoption to diplomacy, one could also make the argument that certain conditions may discourage MFAs from pursuing AI integration or, if they have already started to incorporate it in their work, to slow down the process of AI adoption or even to abandon it. The 'unknown unknowns' discussed in the previous section offer some good examples of AI exit points. If the integrity of the AI system is being systemically compromised (intentionally by other parties by hacking or cyber warfare, or unintentionally due to technical limitations), then the value of AI advice declines sharply, forcing MFAs to completely rethink the feasibility of its AI strategy. A possible regulatory race to the bottom may also gradually erode public and by extension political support for AI transformation on account of its potential for being misused and abused, especially in relation to human rights. As AI continues to develop and become more sophisticated, concerns over the lack of its explainability could accumulate and backfire as the making of decisions without a clear understanding of the underlying rationale would likely be deemed as politically unfeasible. The opportunity costs of AI adoption should not be easily dismissed either. As a new technology, AI is likely to require substantial budgets, at least in the early stages, to cover the cost of research, implementation, and training. The costs may prove daunting for many MFAs, especially for the foreign services of small countries, which may need to see clear return value on their investments before agreeing to embark on a project that may take valuable resources from other policy priorities.

From a broader perspective, AI exit points may be determined by developments in four key areas. The first one relates to the sustainability of AI technological innovation and the confidence that policy makers would continue to have in the contributions that AI has promised to make to the diplomatic field. This confidence cannot be taken for granted. As Bostrom has shown, the quest for artificial intelligence has travelled through multiple "seasons of hope and despair".[66] The early attempts in the 1950s at the Dartmouth College sought to provide a proof of concept for AI by demonstrating that machines were able to perform complicated logical tests. Following a period of stagnation, another burst of innovative thinking took place in early 1970s, which showed that logical reasoning could be integrated with perception and used to control physical activity. However, difficulties in scaling up AI findings soon led to an "AI winter" of declining funding and increased scepticism. A new springtime arrived with the launch of the Fifth Generation Computer System Project by Japan in early 1980s, which led to the proliferation of expert systems as new tools of AI-supported decision-making. After another period of relative stagnation, the introduction of neural networks and deep learning in late 1990s has generated a new wave of interest in AI and growing optimism in the possibility of applying it to a wide range of activities, including diplomacy. The key question on the mind of policy-makers at the moment is whether AI would be able to deliver on its promises instead of entering another season of scepticism and stagnation. If AI would be able to demonstrate value in a consistent manner by providing reliable assistance in areas of diplomatic interest such as in consular services and international negotiations, as suggested above, then the future of AI in diplomacy should look bright. If, on the other hand, the ratio between costs and contributions of AI applications to diplomatic work would stay high, then the appetite for AI integration would likely decline.

66 Bostrom, *Superintelligence : Paths, Dangers, Strategies,* 6–11.

Second, as the level of collaboration between defence and civilian AI sectors continues to gain strength,[67] it becomes debatable whether AI will continue to be seen by MFAs as an opportunity, rather than as a liability for diplomatic engagement. Horowitz argues that AI applications have the potential to shape how countries fight in several ways ranging from increasing the speed with which countries can fight to developing new concepts of operation that could influence how militaries plan operations and how they organize themselves through layers of algorithms that work together to help manage complex operations.[68] Interestingly, Scharre thinks that the ongoing AI arms race favours the development of a new type of security dilemma in which the real danger is not that a country would fall behind its competitors in AI, but that the very perception of the race would prompt everyone to rush to secure the first-mover advantage by pre-emptively deploying unsafe AI systems.[69] Unless clear and feasible dual-use rules for AI technology are developed, the concern for maintaining a competitive advantage in AI technology for national security reasons could clash with the argument that civilian and defence-oriented AI applications are better off developing autonomously of each other. In other words, an accelerated AI arms race may favour a shift in policy-makers' thinking towards imposing stricter scrutiny and possible restrictions on civilian applications of AI, including in diplomacy. This may not necessarily block AI applications to diplomacy, but it would likely slow it down to alleviate concerns over the possibility of AI technological diffusion leading to potential complications for national security.

Third, ethical considerations are likely to play an important role in the future development of AI in general, and in diplomacy in particular. From an ethical perspective, two issues are of particular relevance. First, the level of control diplomats may exert over AI-enabled platforms invites questions regarding decision-making accountability. Over-reliance on AI systems could have damaging consequences in times of diplomatic crises or international negotiations so the issue of human agency ought to be carefully considered. Second, the AI capacity to enable high levels of social control at reasonable costs is real and potentially devastating but its consequences. As Wright points out, by allowing governments to monitor, understand, and control their citizens far more closely than ever before, AI creates conducive conditions for the rise of the digital authoritarian state, possibly even as a political alternative to liberal democratic systems.[70] Trade agreements involving AI technology may thus well require an extra level of legal scrutiny in the same way that arms exports to conflict zones are generally restricted today.

Fourth, the marriage between AI and institutional surveillance has also practical implications for the conduct of diplomacy. From the MFA perspective, the issue bears two possible implications. Internally, it may create an environment in which diplomats may feel less protected in their work as constant AI surveillance would make

67 Kirsten Gronlund, "State of AI: Artificial Intelligence, the Military and Increasingly Autonomous Weapons," Future of Life Institute, 2019, https://futureoflife.org/2019/05/09/state-of-ai/?cn-reloaded=1.

68 Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," 47. See also CRS, "Artificial Intelligence and National Security," 2019, https://crsreports.congress.gov.

69 Paul Scharre, "Killer Apps: The Real Dangers of an AI Arms Race," *Foreign Affairs,* 2019, https://www.foreignaffairs.com/articles/2019-04-16/killer-apps.

70 Nicholas Wright, "How Artificial Intelligence Will Reshape the Global Order," *Foreign Affairs,* 2018, https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order.

them particularly vulnerable to undue political pressure. Externally, it may set in motion a logic of entrenched diplomatic distrust as any embassy or delegation in the world would be theoretically at risk of having their activities clandestinely scrutinised in a such intimate manner that would degrade diplomatic activities to a level close to redundancy. These two factors may combine to generate a dynamic in which diplomats, instead of moving to embrace AI in their work, would actively seek to develop tactics and techniques that can help them contain or even sabotage further adoption of AI systems in their work.

3.3. Key Points

- To unpack AI's potential for diplomacy, we need to map the areas of diplomatic activity in which AI could make a difference, to explore the nature of AI contributions, and to understand the risks they may entail for diplomatic work.
- The degree of decisional structuredness, the method of decision analysis and the type of embodied knowledge have strong analytical value as conceptual benchmarks for assessing AI entry points in diplomacy.
- Decisional structuredness speaks to the issue of the practical relevance of AI in a particular diplomatic field, decision analysis zooms in on the practical contribution that AI can make to diplomatic decision-making processes, while knowledge embodiment calls attention to the question of practical feasibility.
- Structured decisions, descriptive analytics and equipping embodied knowledge are the most likely entry points for AI adoption in the diplomatic field; they apply well to consular services and international negotiations.
- Degrees of sustainability of AI technological innovation, perceptions of AI liability for diplomatic engagement, ethical reflections on matters concerning human control and AI (geo)political implications as well as practical considerations related to surveillance, constitute potential exit points that may lead MFAs to limit, postpone or abandon AI adoption.

4. Designing AI for diplomacy: The TIID Framework

> “AI is akin to building a rocket ship. You need a huge engine and a lot of fuel. The rocket engine is the learning algorithms, but the fuel is the huge amounts of data we can feed to these algorithms” — Andrew Ng (VP & Chief Scientist of Baidu)

Paraphrasing Ng, one could argue that the configuration of AI entry and exit points represents the engine that propels the next generation of the digital diplomacy ‘rocket ship’. However, the fuel that feeds the ‘rocket’ is made of specific AI models designed, built and deployed in support of diplomatic tasks, objectives and strategies. Surprisingly, the ‘fuel’ dimension has been barely noticed in current debates on the role of AI in diplomacy, although not necessarily for the wrong reasons. AI modelling for diplomacy faces natural obstacles related to the continuous evolution of AI technology, the difficultly of converting AI models developed for other areas to the rather unique field of diplomacy, or the thorny issue of MFAs struggling to develop internal capacity and attract the pool of talent that would allow them to take advantage of AI opportunities.

In an effort to bridge this gap, this section advances the **TIID framework** *(task, innovation, integration, deployment)* for designing and integrating AI solutions into diplomatic activity and functions. The framework proposes a particular sequence of model design, which begins with an examination of the specific profile of the diplomatic task that is expected to be improved, continues with an evaluation of the type of innovation required for restructuring the service, a discussion of the level of integration of the physical and digital dimensions of the service, and concludes with an examination of the availability and suitability of the existing institutional configuration. The logic behind the sequence is that AI integration should be service- rather than technology-focused so that the risk of designing and delivering AI-driven services of questionable practical value could be suitably minimised.

The TIID framework can be explained as follows:

Task – the decision to support AI integration should be preceded by a discussion of two related issues: first is the AI solution expected to augment the capacity of humans to deliver the task or to replace them completely? As mentioned above, AI could work well in the case of repetitive and routinised tasks, but less so when human judgement is needed to assess the strategic value of the decision. Second, should the AI solution involve a data-informed or data-driven approach to delivering the service? In the first case, the human would remain in control of how the data is converted into action, while in the latter case the action is largely shaped and implemented by the AI. Chatbots used in consular services are data-driven solutions that replace the work of humans. By contrast, an AI assistant will augment the capacity of a diplomat to conduct negotiations by providing him/her with useful and relevant information in real-time. A more nuanced position could emerge when a data-driven network of AI assistants would be monitored and supervised by a human such as in the potential case of an online public diplomacy campaign.

Innovation – Saving time, resources and costs are the usual drivers of innovation, but problem framing largely informs the scope and value of the innovation response. If, for instance, the issue of crisis management is defined around the idea of developing a tool that can provide more accurate and reliable information to the members of the embassy in times of a crisis, the response will aim to design an AI solution capable of collecting, processing and interpreting vast amount of information in a short period of time. If, on the other hand, the same issue will be defined as a priority matter for the embassy to reach out to nationals, the response will focus on developing an AI tool that can communicate with them in an effective fashion. The analytical component of the innovation needs also to be carefully assessed. Is the AI solution expected to primary provide descriptive, diagnostic, predictive insight to solving the task at hand? In the first example above, descriptive analytics would be most useful, while in the second case diagnostic insight could reveal the most suitable method for engaging with the nationals. Finally, developing a good understanding of one's blind spots is also critical for recognizing the potential limitations of the innovation solution. An AI tool designed for engaging with nationals in times of a terrorist attack might be completely useless during an earthquake-induced crisis, when the entire communication infrastructure might collapse.

Integration – The innovation solution needs to provide a high return value, especially when the physical-digital-physical loop is considered. The *physical-to-digital* part of the integration sequence involves the creation of a digital record based on the information captured from the physical world. The second step, *digital-to-digital,* refers to the use of algorithms for recognizing meaningful signals and patterns in the digital record. The last component, *digital-to-physical,* is about the use of digital insights for driving action back in the physical world via real-time and informed decision-making. The return value of the integration process is therefore defined by the ability of the AI solution to collect relevant information effectively, process it insightfully, and to feed it back into decision-making. An AI assistant designed to counter disinformation will not be useful, for instance, if it is not able to collect data from a variety of sources, identify the most relevant sources and patterns of disinformation and then design the most effective strategies for countering them.

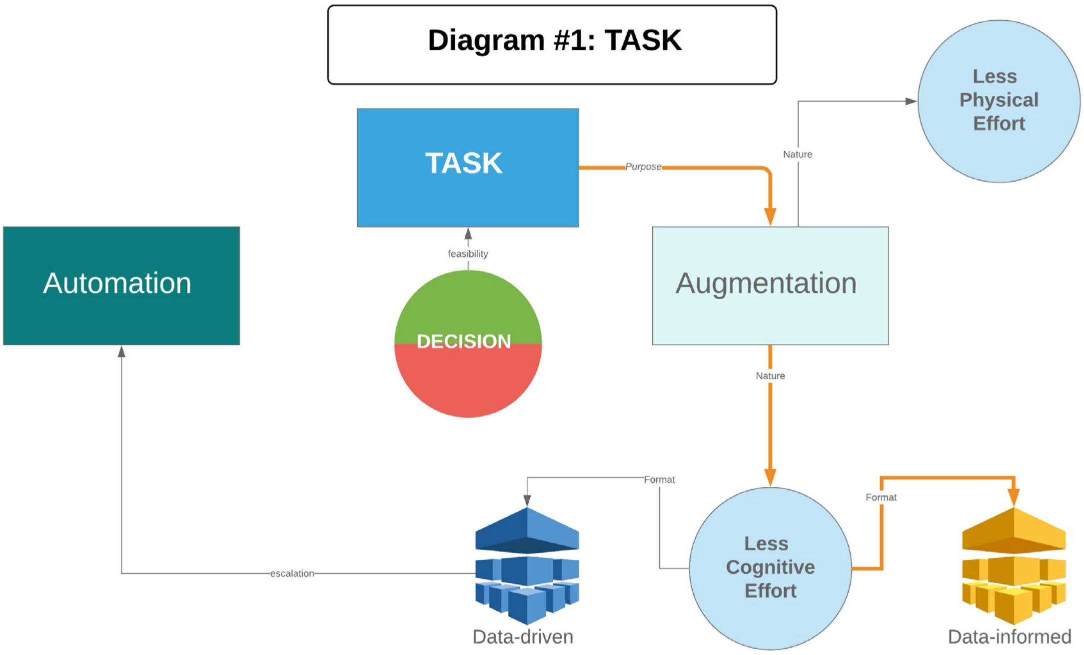
Deployment – Once the decision is taken about how to redesign the service, using tailored innovation and physical-digital integration, the question regarding the suitability of the MFA institutional capacity needs to be considered as well for the deployment of the AI model. While issues concerning costs, level of expertise, training and scale are likely to dominate considerations on the MFA side,[71] one should nevertheless bear in mind that it is not only the machine learning software that matters, but also its potential for integration with other emergent technologies such as 3D printing, Internet of Things (IoT), Virtual and Augmented Reality or Cloud technologies. An AI solution designed for maximising the impact of a development project could well end up recommending a digital-to-physical outcome involving 3D Printing. Similarly, an AI project seeking to streamline the management of embassy logistics could favour solutions involving IoT sensors and VR representations. It is therefore important when modelling AI for diplomacy to stay abreast of the developments in connected areas of digital transformation.

4.1. Case Study: Diaspora Engagement

Let us assume that the MFA of country X decides to make diaspora engagement a policy priority and would like to design and deploy an AI solution in support of this objective. The TIID framework offers a conceptual roadmap for understanding the steps to be taken for modelling the AI solution.

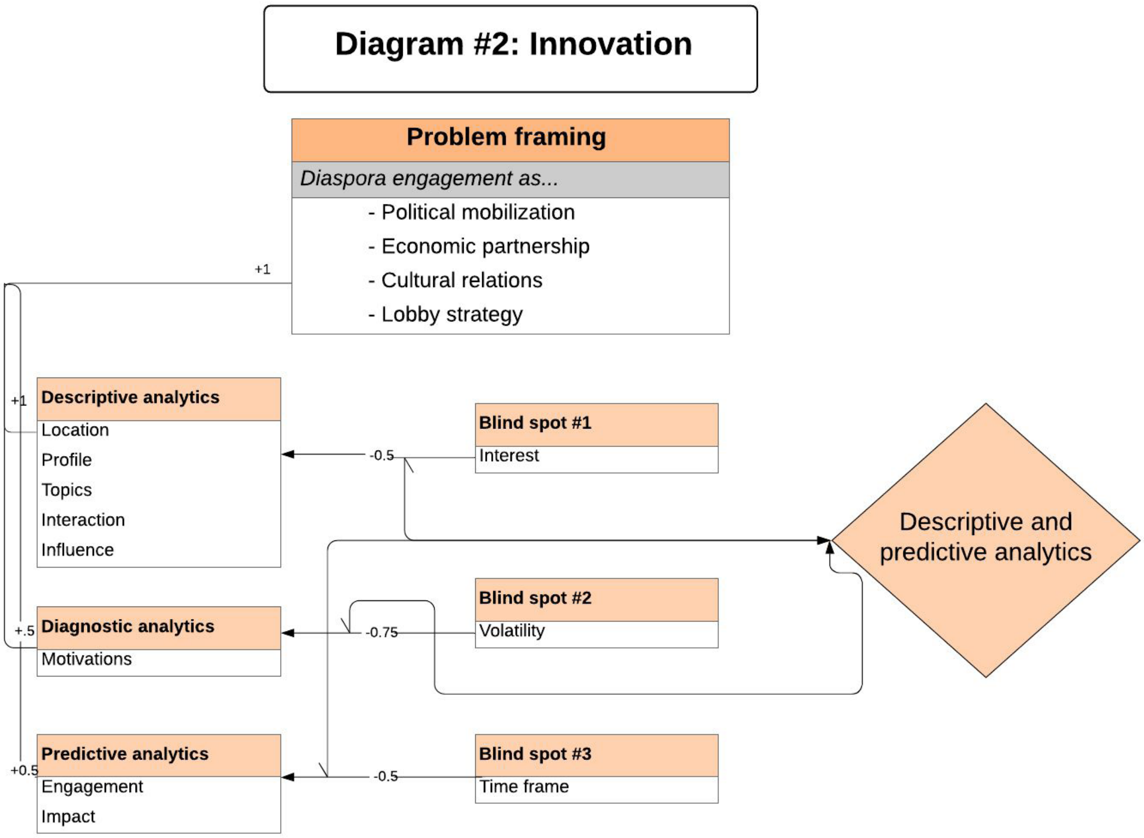
The first step is to define the profile of the service that can help the MFA boost its level of engagement with the diaspora community. The option of replacing humans with a fully automated service might not be feasible in this case, at least in the initial stage, primarily because the objective requires relationship building and a careful balance of the interests of both sides. Using AI to augment diplomats' ability to engage with diaspora communities invites questions about the nature (reducing physical vs cognitive effort) and format of augmentation (data-inspired vs data-driven). It is reasonable to assume in this case that augmentation is expected to improve diplomats' cognitive capacity to better understand the mosaic of profiles and preferences of the diaspora communities and a data-informed approach could generate such insight most effectively, at least in the initial stage. A data-driven approach could become relevant if the volume of data significantly increases over time and challenges the cognitive capacity of the diplomat to process it in an effective manner. The sequence of yellow arrows in Diagram #1 suggests a possible pathway for AI modelling in the first stage: service design pursuing cognitive augmentation using a data informed approach.

71 For a comparison of the Top 8 Artificial Intelligence Software, see https://www.datamation.com/artificial-intelligence/top-artificial-intelligence-software.html



The second step is to identify the ingredients of a potential innovation that would make possible for the MFA to follow and implement the pathway identified in the first stage. Problem framing is the starting point of the discussion since the issue of diaspora engagement could be defined in multiple ways (see Diagram #2). Depending on the MFA's policy agenda, diaspora engagement could focus on political mobilization (such as before an election in the home country), economic partnership (to increase diaspora investments in the home country), cultural relations (to facilitate cultural ties between the diaspora community and the home country) or as a lobby strategy (to mobilize diaspora to lobby the local government in support of a policy of interest to the home country). Let us assume, for the sake of the argument, that the third option is the one favoured by the MFAs as indicated by the +1 score.[72] The question turns then to the type of analytics that could most effectively address the problem as framed above. Descriptive analytics is clearly a must (+1 score) as it helps reveal important information about the members of the diaspora including location, socio-demographic profile, favoured topics of conversations, levels of online engagement and influence.

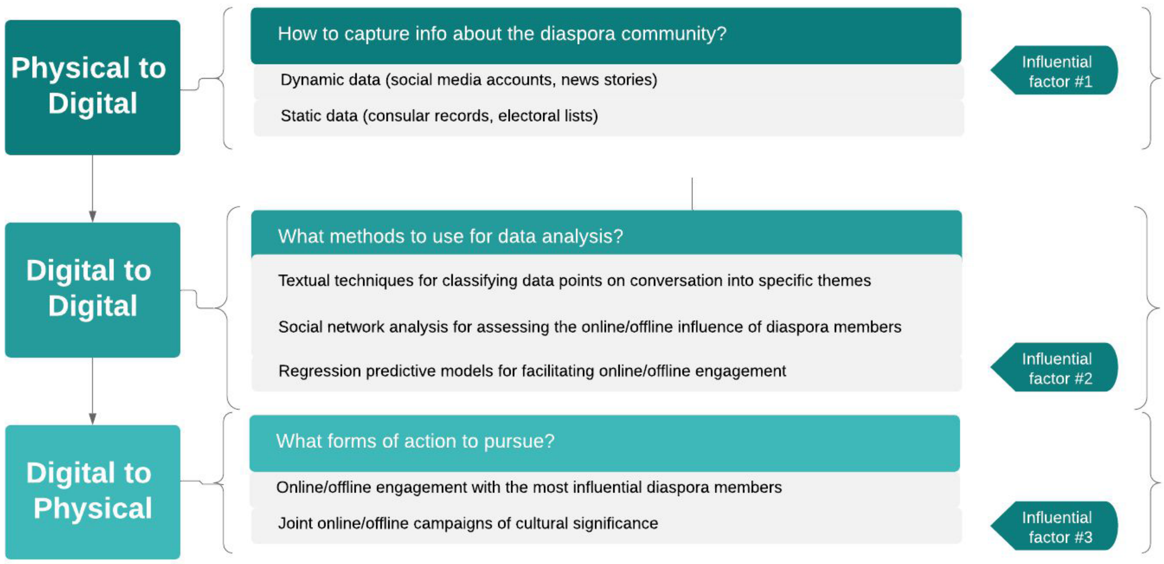
72 The score range in this example is between -1 (low) and +1 (high).



Lack of interest in collaborating with the MFA could be one of the blind spots of descriptive analytics (Blind Spot #1). Diagnostic analytics could investigate the possible motivations of diaspora communities of engaging in cultural relations, but the added value of this effort is likely modest, especially considering that such motivations are usually volatile (Blind Spot #2). Lastly, predictive analytics could offer insight into the conditions for online engagement and likely impact, as long as the time scale of such predictions is duly acknowledged (Blind Spot #3). In short, descriptive and predictive analytics reveal themselves as the most conducive approaches to developing an AI solution for engaging diaspora for cultural purposes.

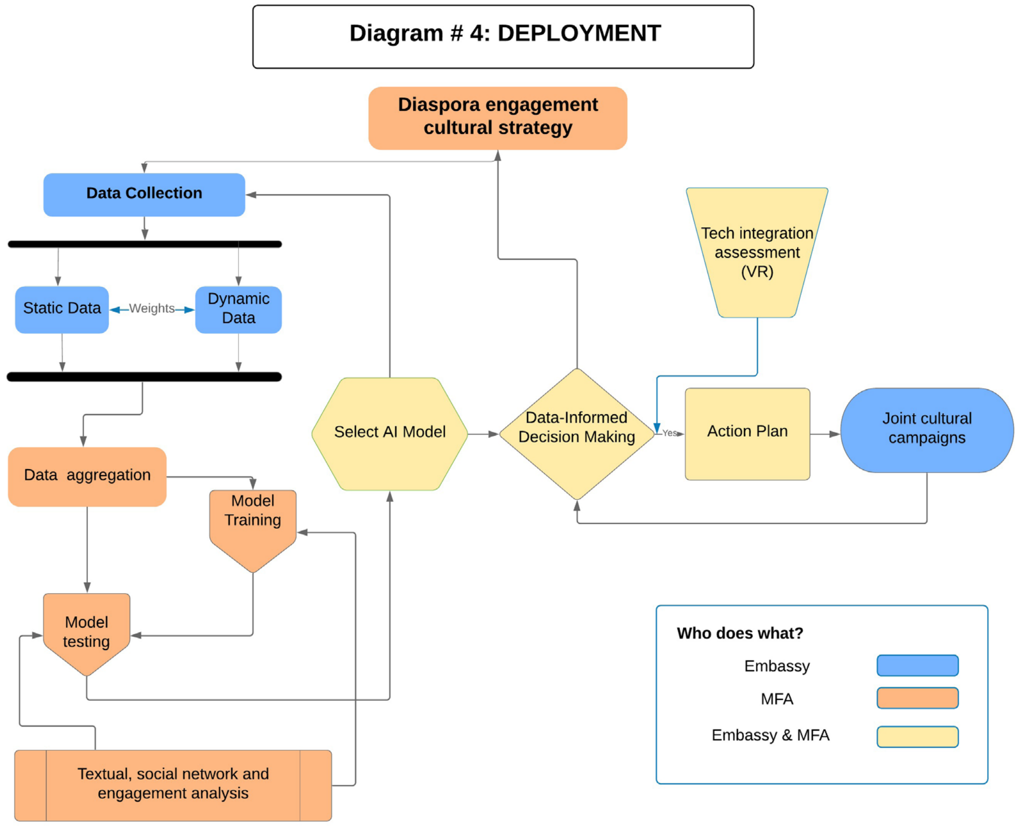
The third step refers to the physical-digital-physical integration loop by which physical and digital data are combined to provide support for conducting descriptive and predictive analysis as discussed in the second stage. Provided that local privacy regulations are strictly followed, electoral registration lists and consular records could be aggregated to build a large dataset with relevant information regarding the socio-demographic profile of the diaspora community. Unfortunately, this data set is relatively static that is, the information is updated slowly, with every new election or passport renewal respectively. Data extracted from the social media accounts of the local embassy and MFA or from news stories published by the local press is much more dynamic and provide crucial information about the level of engagement, topics of interest and patterns of communication and influence of the target diaspora community. It is therefore reasonable to argue that the set of dynamic data is particularly important for the success of the project (influential factor #1), so weights between the two sets of data should be considered. The data so gathered could be then analysed for identifying underlying themes of cultural interest, mapping the network of online and offline

Diagram #3: Physical/Digital Integration



influencers, and for using regression techniques to identify suitable approaches for stimulating engagement on a bilateral basis (influential factor #2). The insight gained from data analysis could be then converted into a plan of action to include proactive online and offline engagement with influential members of the diaspora community and especially joint cultural campaigns between the embassy/MFA and the diaspora community (influential factor #3).

The fourth step involves the building of the AI model taking into consideration all factors discussed in stage 1-3. As indicated in Diagram #4, this will start with a process of aggregation of the data gathered by the embassy from static and dynamic sources. The dataset so generated would then be split into two subsets



(usually 70% training, 30% testing) to be used for training and testing models created with AI algorithms. After running and fine-tuning competing models of textual, social network and engagement analysis, an optimal AI model would be then selected to offer insight to decision-making. The model should be able to indicate the set of themes, the network of influencers, and the format of engagement that could most effectively deliver cultural campaigns jointly organised by the embassy and the diaspora community. The resulting action plan would also include an assessment of the feasibility of integrating other technologies (perhaps VR in this case) for maximising the impact of the cultural campaign. Coordination between the MFA and the embassy would be essential for the success of the project, partly because the embassy would likely lack the necessary resources for designing and deploying the AI model and partly because the process and results would be relevant to other embassies in the MFA network.

To conclude, diaspora engagement could be traditionally pursued by the MFA with existing means and resources with no AI support. What AI brings new is cognitive augmentation, improved effectiveness and speed. It basically allows the MFA and its embassies to 'do more with less' as long as they have taken steps to carefully develop and optimise their technical and human capacity to deploy AI in support of their objectives. It is also important to note that AI integration into diplomatic work cannot proceed in a conceptual vacuum, with no clear direction about what the objective is (task improvement), how to accomplish it (innovation), with what resources (physical/digital integration) and in what institutional configuration (deployment). The TIID framework seeks to bridge this gap by providing a conceptual roadmap for designing, delivering and deploying AI solutions of practical value for diplomatic work.

4.2. Key Points

- The configuration of AI entry and exit points represents the engine that propels the next generation of the digital diplomacy 'rocket ship'. However, the fuel that feeds the 'rocket' is made of specific AI models designed, built and deployed in support of diplomatic tasks, objectives and strategies.
- Diplomatic functions could still be traditionally pursued by the MFA with existing means and resources with no AI support. The new benefits that AI could bring are cognitive augmentation, improved effectiveness and speed.
- AI integration should be service- rather than technology-focused so that the risk of designing and delivering AI-driven services of questionable practical value could be suitably minimised.
- Ministries of Foreign Affairs could deploy the TIID framework as a conceptual roadmap for designing, delivering and deploying AI solutions in diplomacy that combines considerations about what the objective is (service improvement), how to accomplish it (innovation), with what resources (physical/digital integration) and in what institutional configuration (deployment).

Bibliography

Allen, Greg, and Taniel Chan. "Artificial Intelligence and National Security," 2017. www.belfercenter.org.

Allen, Gregory C. "Project Maven Brings AI to the Fight against ISIS - Bulletin of the Atomic Scientists." The Bulletin of the Atomic Scientists, 2017. https://thebulletin.org/2017/12/project-maven-brings-ai-to-the-fight-against-isis/.

Anderson, Paul A, and Stuart J Thorson. "Artificial Intelligence Based Simulations of Foreign Policy Decision Making'." *Behavioral Science* 27, no. 2 (1982): 176–93. https://onlinelibrary.wiley.com/doi/pdf/10.1002/bs.3830270211.

Bauer, MW, and A Suerdem. "Relating 'Science Culture' and Innovation," 2016. https://www.oecd.org/sti/097 - OECD Paper attitudes and innovation_v4.0_MB.pdf.

Bjola, Corneliu. "How Should Governments Respond to Disasters in the Digital Age? | RUSI." The Royal United Services Institute (RUSI), 2017. https://rusi.org/commentary/how-should-governments-respond-disasters-digital-age.

Bjola, Corneliu, Jennifer Cassidy, and Ilan Manor. "Public Diplomacy in the Digital Age." *The Hague Journal of Diplomacy* 14, no. 1–2 (April 22, 2019): 83–101. https://doi.org/10.1163/1871191X-14011032.

Bostrom, Nick. *Superintelligence : Paths, Dangers, Strategies.* First edit. Oxford: Oxford University Press, 2014.

Boulanin, Vincent. "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk." SIPRI, 2019. https://www.sipri.org/publications/2019/other-publications/impact-artificial-intelligence-strategic-stability-and-nuclear-risk.

Boyd, Joshua. "The Facebook Algorithm Explained and How to Work It | Brandwatch." BrandWatch, 2019. https://www.brandwatch.com/blog/the-facebook-algorithm-explained/.

Bremer, Stuart A. "Computer Modeling in Global and International Relations: The State of the Art By My Definition, a Global Model Is a Multisector, Multi-Nation Computer Model of Long-Term Global Dynamics, While an Inter-National Relations (Hereafter IR) Model Is a Multination Model of in-Ternational Political Interactions." Accessed May 30, 2019. https://journals.sagepub.com/doi/pdf/10.1177/089443938900700406.

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, et al. "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," February 20, 2018. http://arxiv.org/abs/1802.07228.

Burstein, F., & Holsapple, C. *Handbook on Decision Support Systems 1: Basic Themes.* Berlin: Springer-Verlag, 2008.

CRS. "Artificial Intelligence and National Security," 2019. https://crsreports.congress.gov.

D'souza, Rhett. "Symbolic AI v/s Non-Symbolic AI, and Everything in Between?" OpenDeepTech, 2019. http://opendeeptech.com/symbolic-ai-v-s-non-symbolic-ai-and-everything-in-between.

Dhande, Meenal. "What Is the Difference between AI, Machine Learning, and Deep Learning." Geospatial World, 2017. https://www.geospatialworld.net/blogs/difference-between-ai-machine-learning-and-deep-learning/.

Dickow, Marcel, and Daniel Jacob. "The Global Debate on the Future of Artificial Intelligence. The Need for International Regulation and Opportunities for German Foreign Policy," 2018. https://www.swp-berlin.org/fileadmin/contents/products/comments/2018C23_dkw_job.pdf.

DiploFoundation. "Mapping the Challenges and Opportunities of Artificial Intelligence for the Conduct of Diplomacy," 2019. https://www.diplomacy.edu/AI.

Doran, Derek, Sarah Schulz, and Tarek R Besold. "What Does Explainable AI Really Mean? A New Conceptualization of Perspectives." Accessed June 3, 2019. http://amueller.github.

Drezner, Daniel W. "Technological Change and International Relations." *International Relations,* March 20, 2019, 004711781983462. https://doi.org/10.1177/0047117819834629.

Drum, Kevin. "Welcome to the Digital Revolution." *Foreign Affairs,* 2018. https://www.foreignaffairs.com/articles/world/2018-06-14/tech-world.

Duan, Yanqing, John S. Edwards, and Yogesh K Dwivedi. "Artificial Intelligence for Decision Making in the Era of Big Data – Evolution, Challenges and Research Agenda." *International Journal of Information Management* 48 (October 1, 2019): 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021.

Edwards, J S, Y Duan, and P C Robins. "An Analysis of Expert Systems for Business Decision Making at Different Levels and in Different Roles." *European Journal of Information Systems* 9, no. 1 (March 19, 2000): 36–46. https://doi.org/10.1057/palgrave.ejis.3000344.

Edwards, John S. "Expert Systems in Management and Administration - Are They Really Different from Decision Support Systems?" *European Journal of Operational Research* 61, no. 1–2 (August 25, 1992): 114–21. https://doi.org/10.1016/0377-2217(92)90273-C.

Forbus, Kenneth D. “AI and Cognitive Science: The Past and Next 30 Years.” *Topics in Cognitive Science* 2, no. 3 (March 8, 2010): 345–56. https://doi.org/10.1111/j.1756-8765.2010.01083.x.

Gan, Chuang. “The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision.” In *Iclr 2019,* 2019.

Garnelo, Marta, and Murray Shanahan. “Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations.” *Current Opinion in Behavioral Sciences* 29 (October 1, 2019): 17–23. https://doi.org/10.1016/J.COBEHA.2018.12.010.

Gartner. “Magic Quadrant for BI Platforms. Analytics Value Escalator.,” 2012.

Geist, Edward, and Andrew Lohn. *How Might Artificial Intelligence Affect the Risk of Nuclear War?* RAND Corporation, 2018. https://doi.org/10.7249/PE296.

Godin, Benoit, and Yves Gingras. “What Is Scientific and Technological Culture and How Is It Measured? A Multidimensional Model.” *Public Understand. Sci.* Vol. 9, 2000. https://archipel.uqam.ca/491/1/Sci_cult_Pub_Und_sc.PDF.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. “Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts.” *Journal of Artificial Intelligence Research* 62 (July 31, 2018): 729–54. https://doi.org/10.1613/jair.1.11222.

Gronlund, Kirsten. “State of AI: Artificial Intelligence, the Military and Increasingly Autonomous Weapons.” Future of Life Institute, 2019. https://futureoflife.org/2019/05/09/state-of-ai/?cn-reloaded=1.

High-Level Expert Group on Artificial Intelligence. “A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines,” 2019. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

Hintze, Arend. “Understanding the Four Types of AI, from Reactive Robots to Self-Aware Being.” The Conversation, 2016. https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616.

Hofstede, Geert, Bram Neuijen, Denise Daval Ohayv, and Geert Sanders. “Measuring Organizational Cultures: A Qualitative and Quantitative Study Across Twenty Cases.” *Administrative Science Quarterly* 35, no. 2 (June 1990): 286. https://doi.org/10.2307/2393392.

Horowitz, Michael C. “Artificial Intelligence, International Competition, and the Balance of Power.” *Texas National Security Review,* 2018.

Horowitz, Michael C, Gregory C Allen, Edoardo Saravalle, Anthony Cho, Kara Frederick, and Paul Scharre. "Artificial Intelligence and International Security." Washington D.C., 2018. https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security.

Landgrebe, J., and B. Smith. "There Is No General AI: Why Turing Machines Cannot Pass the Turing Test," June 9, 2019. http://arxiv.org/abs/1906.05833.

Makridakis, Spyros. "The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms." *Futures* 90 (June 1, 2017): 46–60. https://doi.org/10.1016/J.FUTURES.2017.03.006.

Marr, Bernard. "10 Amazing Examples Of How Deep Learning AI Is Used In Practice?" Forbes, 2018. https://www.forbes.com/sites/bernardmarr/2018/08/20/10-amazing-examples-of-how-deep-learning-ai-is-used-in-practice/#4ee2c912f98a.

McCarthy, John. "What Is AI? / Basic Questions." 2011. Accessed May 22, 2019. http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.

Michael E. Bratman. *Intention, Plans, and Practical Reason.* Cambridge University Press, 1999.

Mintz, Alex, and Karl DeRouen. *Understanding Foreign Policy: Decision Making. Understanding Foreign Policy: Decision Making.* Cambridge University Press, 2010. https://doi.org/10.1017/CBO9780511757761.

Nilsson, Nils J. *The Quest for Artificial Intelligence: A History of Ideas and Achievements.* Cambridge: Cambridge University Press, 2010. https://doi.org/10.1017/CBO9780511819346.

Oxford Insights. "Government Artificial Intelligence 2019 Readiness Index," 2019. https://ai4d.ai/wp-content/uploads/2019/05/ai-gov-readiness-report_v08.pdf.

Pee, L. G., Shan L. Pan, and Lili Cui. "Artificial Intelligence in Healthcare Robots: A Social Informatics Study of Knowledge Embodiment." *Journal of the Association for Information Science and Technology* 70, no. 4 (April 1, 2019): 351–69. https://doi.org/10.1002/asi.24145.

Pomerol, Jean-Charles. "Artificial Intelligence and Human Decision Making." *European Journal of Operational Research* 99, no. 1 (May 16, 1997): 3–25. https://doi.org/10.1016/S0377-2217(96)00378-5.

Prakash, Abishur. "Algorithmic Foreign Policy." Scientific American Blog Network, 2019. https://blogs.scientificamerican.com/observations/algorithmic-foreign-policy/.

Roy, Asim. "A Theory of the Brain - the Brain Uses Both Distributed and Localist (Symbolic) Representation." In *The 2011 International Joint Conference on Neural Networks,* 215–21. IEEE, 2011. https://doi.org/10.1109/IJCNN.2011.6033224.

Rumsfeld, Donald H. "Defense.Gov Transcript: DoD News Briefing - Secretary Rumsfeld and Gen. Myers." U.S.Department of Defense, 2002. https://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence A Modern Approach Third Edition. Pearson,* 2010. https://doi.org/10.1017/S0269888900007724.

Santos, Maximiliano Ribeiro Aquino. "Cognitive Trading Using Watson." IBM blog, 2018. https://www.ibm.com/cloud/blog/cognitive-trading-using-watson.

Savov, Vlad. "Amazon Preparing a Wearable That 'Reads Human Emotions,' Says Report - The Verge." The Verge, 2019. https://www.theverge.com/circuitbreaker/2019/5/23/18636839/amazon-wearable-emotions-report.

Scharre, Paul. "Killer Apps: The Real Dangers of an AI Arms Race." *Foreign Affairs,* 2019. https://www.foreignaffairs.com/articles/2019-04-16/killer-apps.

Scott, Ben, Stefan Heumann, and Philippe Lorenz. "Artificial Intelligence and Foreign Policy," 2018. https://www.stiftung-nv.de/sites/default/files/ai_foreign_policy.pdf.

Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, et al. "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play." *Science (New York, N.Y.)* 362, no. 6419 (December 7, 2018): 1140–44. https://doi.org/10.1126/science.aar6404.

Stone, Peter et al. "Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, 52.," 2016. https://ai100.stanford.edu.

Trent, Stoney, and Scott Lathrop. "A Primer on Artificial Intelligence for Military Leaders." *Small Wars Journal,* 2019.https://smallwarsjournal.com/jrnl/art/primer-artificial-intelligence-military-leaders.

Ullman, Shimon. "Using Neuroscience to Develop Artificial Intelligence." *Science (New York, N.Y.)* 363, no. 6428 (February 15, 2019): 692–93. https://doi.org/10.1126/science.aau6595.

White House. "Artificial Intelligence, Automation, and the Economy," 2016. https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED AI Economy Report.pdf.

Wright, Nicholas. "How Artificial Intelligence Will Reshape the Global Order." *Foreign Affairs,* 2018. https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order.