

!#\$%&'()\*+,-./0+\*\$1(\$0&(2.'3"0'+41.(  
56(!%\$0'7"%8\$-+"0(+0(9.4\$0"0  
"#\$%&%\$'()\*+,-./+,\$01%\$('02!\$3&  
4%#)05+%062!7\$.6'31!81\$%3'39!\*+%!:\$0\$!1)1%0)

!#\$%"& &\*+ &".% ! /\$00"%&"1. !

!#\$%&'()\*+,-./0+\*\$1(\$0&(2.'3"0'+41.(56(!%\$0'7"%8\$-+"0(+0(9.4\$0"0

# **Towards a Socio-technical and Responsible AI Transformation in Lebanon**

Guardrails from Socio-materiality and  
Trustworthy Machine Learning for Data Deserts

*Working Paper*

---

*Fatima K. Abu Salem   Wissam Saade*

**© All Rights Reserved. Beirut, July 2025.**

This Working Paper is published by the Issam Fares Institute for Public Policy & International Affairs (IFI) at the American University of Beirut, made possible (in part) by a grant from Carnegie Corporation of New York, and is available on the following website: <http://www.aub.edu.lb/ifi>.

The views expressed in this document are solely those of the author, and do not necessarily reflect the views of the Issam Fares Institute for Public Policy & International Affairs or that of Carnegie Corporation of New York.

## About the Authors

---

***Fatima K. Abu Salem*** is Professor of Computer Science at the American University of Beirut. She holds an MS in pure mathematics from AUB and a DPhil in Computing from the University of Oxford. Her former research area has been in Computer Algebra, with a focus on developing parallel and cache efficient algorithmic designs for algebraic computations at scale. Her recent research area is in data science with impact, with applications in the social, health, agricultural and environmental sciences. In her work, she incorporates advanced machinery that adheres to trustworthy Machine Learning requirements such as ML for small data, distributional robustness, probabilistic forecasting, and uncertainty quantification. She has spoken multiple times on challenges affecting women in computing and mathematicians in the developing world, and served as secretary for the special activity group on Supercomputing for the Society of Industrial and Applied Mathematics. She currently serves as associate editor for the Journal of Parallel and Distributed Computing, and member of the ACL special interest activity group on Arabic Natural Language Processing.

***Wissam Saade*** is an IFI Associate Fellow and Lecturer of Political Science and History at Saint Joseph University since 2003. He is a regular op-ed writer in leading Lebanese and Arab newspapers. Saade's research interests span medieval and modern political thought, the social and intellectual history of modern revolutions, as well as nationalism and ethnicities in South Asia and the Middle East.

# Table of Contents

<b>Abstract</b>	<b>6</b>
<b>Introduction</b>	<b>6</b>
<b>Collective Intelligence</b>	<b>8</b>
<b>Citizen Science and Data Democratization</b>	<b>9</b>
<b>Decolonial AI</b>	<b>10</b>
<b>Framework for Data Democratization</b>	<b>12</b>
<b>Key Stages of Data Democratization Maturity</b>	<b>12</b>
<b>Assessing Data Democratization Maturity</b>	<b>13</b>
<b>Framework for Responsible AI</b>	<b>14</b>
<b>Ethical AI</b>	<b>14</b>
<b>Accountability Risks</b>	<b>15</b>
<b>Framework for AI Value Creation, Capture, and Destruction</b>	<b>19</b>
<b>Framework for Human Labor in the Age of AI</b>	<b>22</b>
<b>Final Recommendations</b>	<b>23</b>
<b>Socio-materiality</b>	<b>23</b>
<b>Data Democratization</b>	<b>24</b>
<b>Responsible AI</b>	<b>24</b>
<b>Ethical AI</b>	<b>24</b>
<b>Accountability Risks</b>	<b>24</b>
<b>Trustworthy Machine Learning</b>	<b>27</b>
<b>AI Value Creation, Capture, and Destruction</b>	<b>28</b>
<b>Human Labor in the Age of AI</b>	<b>29</b>
<b>Addendum: Analytics for Corruption</b>	<b>29</b>
<b>Conclusion and Final Remarks</b>	<b>31</b>
<b>References</b>	<b>33</b>
<b>Appendices</b>	<b>37</b>

## Abstract

---

As Lebanon embarks on a long overdue data and AI transformation, the daunting challenges and setbacks it has continuously grappled with raise substantially founded concerns around its ability to navigate the associated opportunities and risks, in an increasingly volatile and ambiguous world. In this working paper, we propose forward an ambitious manifesto and a suite of recommendations for complementing the five-year strategy (LEAP) launched by the ministry of state for Technology and AI in Lebanon. Whether among the elites or the general public, the question of artificial intelligence remains, in a country like Lebanon, shrouded in a jungle of myths and confusion. AI is perceived at times as an absolute threat to employment, and at others as a magical solution to all the country's problems. Often, it is reduced to limited applications such as text or image generation. This narrow view limits our ability to grasp the scope and depth of ongoing AI transformations happening worldwide. In Lebanon, as in other countries facing structural collapse, AI is frequently seen as a miraculous shortcut: a means to avoid institutional labor, deep reform, and long-term educational investment. Yet, no technological tool, however advanced, can replace a coherent and dynamic social, political, and economic ecosystem. Such an ecosystem relies on institutions, shared values, and sustained collective effort: elements that technology alone cannot create or sustain. In light of all that, this working paper aims to augment a plethora of channels by which to support the LEAP roadmap using a socio-technical and responsible AI parlance, against a backdrop of talent exodus. How can a data desert such as Lebanon progress robustly and reliably to become an AI hub? What evolving governance and ethical risks need to be navigated with the advent of GenAI and Agentic AI? How do we ensure we are creating value from AI whilst preventing value destruction in other non-AI systems? And how should we revise labor policies, understand human labor in light of required human agency, and safeguard human workers, in the AI age? For all intents and purposes, this manifesto prioritizes the need to render such ambitions credible and attainable, so we can move beyond AI proofs-of-concept towards AI deployment across our public and private sectors, robustly and responsibly.

## Introduction

---

The present working paper is the culmination of its authors' extensive experiences as researchers and practitioners in responsible AI for impact, on one hand, and modern political thought, and the social and intellectual history of modern revolutions, on the other. The amalgam of those two seemingly disparate areas of expertise reflects the increasing need on a global level to conceive of AI solutions that are responsive to societies' practical needs and yet observant of the intricacies of social epistemologies and of higher-order cognitive attributes of individuals as well as collective groups. This working paper strikes a much-needed balance between the socio- and the technical- discourses, calling for co-shaped futures, not technology-first futures, and in doing so, we begin by conjuring up socio-material realities that emphasize the interconnectedness of social phenomena and material elements, technology included, in shaping our human experiences and practices, our understanding of the world, and our actions within it. Under an overarching framework spanning collective intelligence, decolonial AI, citizen science and data democratization, we center our discussion around the four main pillars of data democratization maturity, responsible AI maturity, AI value creation, capture, and destruction, as well as understanding human labor in the age of AI. Guardrails from the emerging technical discipline of trustworthy Machine Learning orient our working paper, as we explore the need to adhere to distributionally robust AI models, algorithmic fairness, explainability and interpretability of AI models, uncertainty quantification of risk, as well as machine learning for small data. Those emerging technical requirements dictate the success of AI systems once take outside the realms of technological labs into those of real life applications and are a prerequisite for societal trust and acceptance of AI systems. With the advent of GenAI and Agentic AI systems, we also revisit various requirements for AI governance and ethical AI and the need to account for newly emerging risks associated with a collective AI-human collaboration and hybrid content generation and role playing. Pushing the limits

of our manifesto beyond requirements for data and AI maturity and capabilities, we overview emerging organizational requirements to create value from AI and maintain (capture) this value, and provide a cautionary tale of emerging risks how, in the quest to create and capture AI value, other non-AI values are destroyed. We also highlight the emerging need to develop agency-driven work environments, where individuals are no longer just executing predefined tasks but are instead responsible for designing systems, exercising judgment, and directing processes, which bears significant implications for our national AI strategy. This shift in theorizing about human labor mandates a thorough rethinking of how we approach both education and labor regulation in the age of AI.

The flow of the current paper mimics how we advocate that strategists, visionaries, and people of expertise in Lebanon, approach the AI transformation journey, and the chronology presented herein delivers a natural progression, slowly but surely, suited for technologically underdeveloped, yet significantly socially complex and intellectually rich societies, such as in Lebanon. The guardrails we develop in this working paper give rise to a set of recommendations that we hope can drive the national AI strategy in Lebanon using state-of-the-art socio-technical metrics and guidelines, as we dedicate a subset of the recommendations for the deep-rooted problem of Corruption. As addendum to our set of recommendations put forth in this working paper, we demonstrate the ethos of our manifesto as it unfolds across several real-world and locally contextualized and localized case studies in research, teaching, and outreach. Conducted by our research teams within AUB and their local and international collaborators, those case studies demonstrate how to use AI responsibly on data that is largely from the region, for the region, and are oriented by elements from participatory data science and trustworthy machine learning, across healthcare, agriculture, and environmental studies, as well as conflict studies and corruption studies, to name a few. This working paper will be presented at the AI retreat of the [AI in Lebanon Conference](#), on July 26, 2025, in Aley, Lebanon.

## Framework for Socio-Materiality

---

Artificial intelligence, far from being a mere technical tool, has become a critical mirror: an opportunity to rethink our knowledge, our reasoning, and ultimately our cognitive and social condition.

For an indeterminate time to come, artificial intelligence stands as the latest culmination of a series of scientific revolutions that, since the rise of modernity in the sixteenth and seventeenth centuries, have been defined by what Gaston Bachelard called an “epistemic rupture”: a deliberate break from the prejudices of everyday common sense (Bachelard, 1938). Just as telescopes and microscopes revealed realities at cosmic and quantum scales that defied our unaided perceptions, and as calculus and thermodynamics forged new conceptual frameworks beyond intuitive grasp, so too do today’s AI models operate on mathematical and statistical principles that lie far outside the realm of human intuition.

Yet this very rupture is paradoxically the source of AI’s power to captivate the human imagination. By functioning according to algorithms and high-dimensional vector spaces, AI behaves in ways that often seem miraculous, leaving behind a gap between how it actually works and how it appears to work. This gap provokes fresh reflections on the nature of mind and meaning, reigniting age-old questions about consciousness, creativity, and the limits of human understanding.

At the same time, AI is not merely a novel scientific instrument, but a transformative force reshaping human intelligence itself. As it augments our cognitive reach (assisting memory through search engines, automating repetitive or calculation-heavy tasks to free us for creative or strategic thought, and personalizing education via adaptive learning systems), it also challenges us to interrogate which skills we choose to preserve or abandon. Debates over cognitive “atrophy” versus skill reorientation remind us that every technological gain carries the risk of new blind spots. Socially and culturally, too, AI forces a reckoning. Interaction with chatbots and virtual assistants reshapes our perception of

social cues; overreliance on algorithmic advice can atrophy critical reasoning even as it cultivates digital literacy, algorithmic awareness, and ethical deliberation. In demanding that we rethink how we learn, relate, create, and understand ourselves, AI both intensifies the rupture with common sense and, in doing so, offers a renewed field for the human spirit's limitless imaginative capacity.

Socio-materiality is a theoretical perspective that explores how social and material aspects are inextricably co-constitutive in shaping practices and outcomes (Vega et al, 2023). It emphasizes that social, cultural, economic, and political forces shape AI, and AI in turn reshapes those forces. Through this mutual shaping we caution against assuming that AI will inevitably lead to development or progress, unless we frame AI as a sociotechnical project that must be developed in tandem with local needs, values, and systems.

The implications of socio-materiality for policy are several-fold: AI strategies must be context-sensitive, allowing for co-evolution between AI technologies and local conditions, rather than imposing one-size-fits-all models. AI is not “neutral” or purely technical; national strategies must treat AI as embedded in broader material and social arrangements, including labor, infrastructure, and power dynamics.

AI is not “neutral” or purely technical; national strategies must treat AI as embedded in broader material and social arrangements, including labor, infrastructure, and power dynamics. Hereafter, we work to situate a suite of recommendations under the socio-materiality framework using the three main pillars of collective intelligence, data democratization, and decolonial AI.

## Collective Intelligence

---

The myth of the AI shortcut reflects less faith in technology than a loss of confidence in collective capacities for transformation and emancipation. In a country marked by communal fragmentation and widespread mistrust, AI is sometimes fantasized as an instrument of total surveillance, capable of “seeing everything,” “hearing everything,” and “controlling everything.”

One must not forget this obsession with “surveilling everything” through AI. Shoshana Zuboff, in her book *The Age of Surveillance Capitalism* (Zuboff, 2019), described a system in which personal data is exploited for commercial and control purposes, warning against the risks of dispossession of privacy and the subjugation of individuals to market-driven and predictive logics. However, this vision has sometimes been misunderstood, leading either to excessive fascination with the omnipotence of “total control” or to alarmist catastrophism. Both attitudes are often fueled by admiration for “smart security” models, where cameras, facial recognition, and AI combine to technologically grid bodies and movements.

A more cautious and critical perspective should instead acknowledge that, to escape the dilemma between technology risking total control and technique escaping all control, it is essential to link the question of AI to fundamental societal choices. This implies open debate, promoting citizen participation in technological choices, establishing rigorous regulation and democratic governance, and insisting on the necessity of a clear, transparent, and protective legal framework. Above all, it calls for encouraging a humanistic posture of technological humility: recognizing that AI is a powerful yet limited tool whose purpose must remain subordinated to human values.

To break free from this impasse, we must first acknowledge that the epistemological and cognitive question of artificial intelligence remains fundamentally open. What is the nature of the knowledge produced by AI? Is it knowledge comparable to that of humans? Or is it merely statistical reproduction, an illusion of understanding, an imitation of language, or even an autonomous form of intelligibility? Thus arises the question of the validity, limits, and status of algorithmic knowledge. This inquiry inevitably leads to a broader and older questioning of the very nature of human

understanding. Indeed, the development of AI provokes a dual movement: it challenges human intelligence, forcing it to reflect on its own modes of thought, and it imposes, in a way, the challenge of becoming clearer-sighted, more lucid, more intelligent.

Several lines of research define collective intelligence as the outcome of collaborative efforts where inputs from a group are synergized to achieve results surpassing individual capabilities, a term hitherto known as crowdsourcing. This is distinct from the concept of the wisdom of the crowd, which emerges from aggregating independent judgments of large, distributed individuals, typically through statistical averaging. Whereas the wisdom of the crowd thrives on independence and diversity of opinion, collective intelligence is rooted in interaction, cooperation, and shared goals, drawing on the dynamics of group performance and decision-making (Cui and Yasseri, 2024).

The applications of AI-enhanced collective intelligence via crowdsourcing span a wide array of domains, demonstrating its potential to address complex societal challenges. From supporting community responses to climate change and sustainability, to aiding real-time crisis management and combating misinformation, AI serves as a powerful decision-support tool. Notably, its reach extends to sensitive areas like healthcare and criminal justice, where the stakes are high. Drawing on the Supermind Design database, which catalogs over 850 cases of AI-enhanced collective intelligence, the majority of crowdsourcing applications are found in the public and NGO sectors, followed by high-tech and media industries (Cui and Yasseri, 2024).

For countries like Lebanon, crowdsourcing represents a particularly promising approach to harnessing collective intelligence. As our discussion above suggests, collective intelligence does not necessarily depend on vast amounts of independent data or highly centralized infrastructures; rather, it emerges from the synergy of human collaboration, diverse perspectives, and interactive problem-solving. Crowdsourcing leverages these very principles by mobilizing resources that remain abundant despite operating within resource-constrained settings, resources such as local knowledge, community participation, and distributed effort. By engaging citizens directly in the collection, interpretation, and validation of data, crowdsourcing helps compensate for institutional gaps and limited centralized capacity while fostering ownership and inclusivity.

The dynamics of crowdsourced systems align well with the realities of Lebanon. The absence of robust funding or infrastructure to maintain sophisticated Internet of Things (IoT) networks, for example, does not preclude effective action if people themselves can act as sensors, annotators, and decision-makers. Crowdsourcing provides a low-cost, scalable alternative to the expensive and maintenance-intensive hardware that IoT requires while still generating timely, contextually relevant insights. In situations where deploying and sustaining an IoT ecosystem is financially or logistically unfeasible, crowdsourcing allows countries to tap into the collective capacity of their populations, often producing actionable information more quickly and flexibly.

Importantly, the participatory nature of crowdsourcing aligns with decolonial and democratic values, ensuring that local voices shape the data and decisions that affect them. Rather than imposing external technologies or extractive practices, crowdsourcing embeds AI and collective intelligence within existing social fabrics, respecting local knowledge and enhancing social cohesion.

## **Citizen Science and Data Democratization**

The contemporary use of the term *citizen science* emerges from two distinct epistemological perspectives rooted in their respective disciplinary origins (Haklay et al., 2021). The first perspective, developed by Alan Irwin (Irwin et al., 1994; Irwin, 1995), emphasizes the involvement of citizens as stakeholders in the outcomes of research, particularly in areas like public and environmental health. Irwin positions citizen science “at the point where public participation

and knowledge production, or societal context and epistemology, meet” (Irwin, 2015). He contends that such approaches create opportunities to foster closer connections between the public and science, advancing the notion of an engaged ‘scientific citizenship’ with direct relevance to public policy. The second perspective, articulated by Rick Bonney (Bonney, 1996), highlights the role of volunteers in contributing observational data on the natural world, coordinated by professional scientists. Both conceptualizations and several analogous ones gave rise to the adjacent notion of citizen science data, more commonly known as data democratization.

Data democratization refers to the process of rendering data accessible and easily usable by all members of an organization, regardless of their role or level of expertise. In principle, it dismantles the traditional barriers that confined data access to specialized departments such as information technology (IT) or data science. The caveat is as follows: by empowering a broader spectrum of individuals with the ability to engage with data, organizations can foster a culture of data-driven decision-making that drives both innovation and operational efficiency.

Viewed from another perspective, data democratization also involves engaging users, not only experts, in the discovery, access, and sharing of data, while maintaining compliance and control. This approach aligns closely with the FAIR principles of data management: ensuring that data is Findable, Accessible, Interoperable, and Reusable (Labadie et al., 2020).

A holistic approach to data democratization discussed extensively in (Džanko et al., 2024) integrates a plethora of socio-material components involving technology, culture, education, and governance to build an efficient and sustainable data ecosystem. Technology provides the necessary tools and infrastructure to facilitate easy access to data and to support its central role in organizational decision-making. At the same time, effective governance remains a critical enabler, ensuring that data democratization is both successful and responsible. Successful democratization also depends on cultural dimensions such as leadership, transparency, and collaboration. Strong leadership sets a clear vision for data-driven practices, encouraging employees to integrate data into their daily operations and maximize its organizational value. Transparency in data practices builds trust by making performance and decision-making processes visible, fostering open communication and confidence in data use. Finally, collaboration across various organizational silos promotes innovation and improves decision-making.

We situate several recommendations put forth in this working paper under the Data Maturity Framework using the reverberating theme of Data democratization.

## Decolonial AI

The question of ownership in artificial intelligence lies at the intersection of legal, economic, ethical, and political debates worldwide, and nowhere is this more urgent than in Lebanon, a country facing unique challenges of data scarcity, digital dependency, cultural and structural vulnerability.

Lebanon, with limited domestic AI infrastructure, finds itself reliant on imported models, often trained on datasets that reflect Western values, languages, and power hierarchies. Within this context, we remark the rapidly escalating argument that digital technologies, including AI, reproduce colonial patterns, and invoke the associated notions of *digital-territorial* and *digital-structural coloniality* to illustrate further (Mohamed et al., 2020). Digital spaces, like physical territories, become sites of extraction and exploitation, while socio-cultural and institutional structures continue to reflect colonial power. Theories such as data colonialism and data capitalism highlight how data functions as a resource exploited for economic and social control, imposing dominant ways of thinking and marginalising alternative worldviews. Building on this, the evolving notion of *algorithmic coloniality*, examining how algorithms reinforce inequalities through resource allocation, labour markets, geopolitical dynamics, and ethical debates

(Mohamed et al., 2020). As we consider integrating AI more deeply into society, particularly in contexts like the Lebanese context, it is vital to establish strong guardrails to ensure this transition is responsible and equitable, against this specific backdrop of algorithmic coloniality. Throughout history, societies have organized themselves around dominant and respected sources of “truth”, whether rooted in religion, rationality, or now increasingly in AI (Miller, 2022). These sources have often been invoked to justify who holds power, how resources are allocated, and who governs. By placing blind trust in such systems, one risks relinquishing their responsibility and agency, granting too much power to mechanisms that can enable monopoly, surveillance, corruption, and oppression.

A modern critical practice of AI grounded in decolonial thought advocates cultivating a *double vision*: recognizing and challenging the centers and peripheries of power through reciprocal, reverse tutelage, while also dismantling colonial binaries (Mohamed et al., 2020). This approach questions dominant assumptions about knowledge, on whether it is absolute and fully representable through data, or inherently incomplete and shaped by values and interpretations. Decisions about what constitutes valid knowledge, what data is collected, and what is excluded reflect power dynamics that must be acknowledged. To confront these dynamics, researchers have been proposing various ways of reverse tutelage, where marginalized perspectives inform and reshape knowledge production (Mohamed et al., 2020).

Whilst an exhaustive decolonization of all dimensions of our present manuscript is deeply merited, we contend that this would be beyond the scope and space of the current working paper. Instead, we simulate a typical debate that reflects the concerns of broad segments of Lebanese society, with regards to the readiness of our country to embark on something as ambitious and demanding as an AI transformation.

Decolonizing AI in Lebanon means reclaiming not just datasets, but the very right to define the categories, values, and epistemes that shape knowledge production. In data-rich countries, debates often focus on who owns or can monetize vast national datasets. But in data-poor environments like Lebanon, the issue is not only about ownership; it is fundamentally about who controls the processes of data collection and under what conditions. The urgent question becomes: Who is collecting our data and on whose terms? Who defines the rules, controls access, and safeguards the rights of data subjects?

***This led us to three key questions:***

***First***, Who has the right to compute in Lebanon? Current reliance on foreign cloud services means that Lebanese data is often processed abroad, raising privacy and security concerns, while local users face high costs and unstable access.

***Second***, can we build small, sovereign, solar-powered AI systems? Developing such systems is feasible with advances in micro-model AI and edge computing but requires investment in local capacity-building and hardware adaptation.

***Third***, what options do we have to become less environmentally taxing?

AI is not simply about intelligence; it is about power, in the most literal sense. Many imagine AI as “virtual” or “smart,” disembodied from material reality. But in truth, AI is profoundly material. As Gilbert Simondon, relayed by Deleuze, explains, electricity is a power of transduction, a force of passage and potential structuring (Simondon, 2005). It is a matter that is not given immediately but traverses, affects, and informs. In this vein, AI can be seen as a capture of electricity by logic: a transductive movement between energetic matter and algorithmic structuring. Artificial intelligence is not thought per se; it is the structuring of flow. Electricity becomes diagram, current becomes syntax. The neuron is no longer cellular; it is matricial. In a way, AI does not *think*. It *functions* electrically. It is an electrification of computation, a conversion of voltage into response. Training large language models or computer vision models, which rely inherently on deep neural networks, requires enormous amounts of energy, often consuming megawatts over weeks. Every query you send to an AI model incurs costs, hungry for electricity, cooling, server

uptime, and network bandwidth. Data centers, known as the “brains” of AI, consume more electricity than some small countries.

In Lebanon’s context, this reality confronts a set of harsh material conditions. The country suffers from chronic electricity shortages and an unreliable national grid. The population widely depends on private generators, amidst several crises affecting our reserves of fuel, infrastructure, and energy sovereignty. If AI models are trained in Lebanon, the likely scenario is that we will yet be burning diesel or running polluting private generators, exposing workers in the AI industry to long outages and health risks.

Only energy-secure institutions can reliably use AI tools at scale. In the face of this, one begins to ask: can we shift to a Micro-Model Alternative? Micro-models represent AI architectures drastically smaller than mainstream large models, with millions rather than billions of parameters. They demand less memory, compute power, and electricity, enabling deployment on low-end devices and on-premises hosting rather than on expensive cloud infrastructure. They transform AI from a “big data, big compute” paradigm to one of smart, local, and accessible intelligence, thereby fitting countries with energy scarcity, infrastructural challenges, and economic limitations. This approach promises a more equitable, sustainable path to AI adoption that respects data sovereignty and local realities, and a more realistic scenario for contexts within Lebanon that are marred with unstable internet, limited electricity, and constrained budgets.

Decolonial AI will inform the conceptualisation of several recommendations we will be proposing in this paper.

## Framework for Data Democratization

Data democratization maturity refers to the extent to which an organization has institutionalized open, secure, and responsible access to data across its structure. In turn, Data Democratization Maturity Models (DDM) refer to specialized frameworks that gauge the extent to which an organization is enabling non-technical and technical users alike to access and integrate data in decision-making, while maintaining data quality and ethical standards. Frameworks help organizations phase out their plans and readiness gradually across stages with measurable KPIs.

As data becomes increasingly central to strategies and performances across all organizations, perhaps the most instrumental requirement nowadays is to relinquish all forms of fragmented, ad-hoc data practices and ensure a scalable, and integrated approach to democratization abiding by well-established maturity frameworks. In this working paper, we explore one of the latest such DDMM models by Džanko et al., 2024, according to which we develop several recommendations to ensure a structured path that helps organizations (1) diagnose their current stage in data democratization maturity, (2) identify the most important challenges, and (3) set clear, actionable goals towards a data democratization transformation.

## Key Stages of Data Democratization Maturity

The DDMM framework presented herein defines multiple levels of maturity, each reflecting an organization's progress in democratizing data. Each stage builds on the previous one by integrating tools, skills, policies, and cultural change to promote responsible and widespread data use.

**Unaware:** In this stage, an organization’s data is scattered, inaccessible, and siloed. The organization lacks any formal initiatives to support broad data access, and its decision-making model is often intuition-based or limited to high-level reports.

**Aware:** In this stage, an organization's leadership team acknowledges the need for better data access, and has probably launched initial assessments, but overall strategy and actions remain fragmented. In this stage, some specialized data tools might be deployed but remain poorly understood and insufficiently integrated across various departments.

**Developing:** In this stage, data governance policies begin to emerge. Specific and highly specialized teams or departments are beginning to use data proactively and with intent. Organizations in this stage begin investing in data literacy programs, cataloging initiatives, and obtaining data integration tools. Yet, challenges affecting data infrastructure, cultural barriers, or data security remain significant.

**Established:** In this stage, an organization rolls out across the board policies and frameworks for data sharing. Key stakeholders (e.g., IT, business analysts, data stewards) are engaged in designing and activating metadata, catalogs, access policies, and role-based permissions. Regular training programs are conducted to improve user confidence in the existing data frameworks and to prevent data misuse.

**Optimized:** In this stage, an organization treats data as a strategic asset across all levels of its operations. It maintains seamless mechanisms to access accurate, robustly governed, and timely, relevant data. The organization has fostered a culture of data literacy and trust in the entire process, specifically with regard to accountability mechanisms underpinning all data-driven decision-making. The organization ensures that feedback loops are put in place for the ongoing refinement of tools, policies, and training.

## Assessing Data Democratization Maturity

In light of the above maturity stages, Džanko et al., 2024 outline tangible activities and methods to assess an organization's current state in the data democratization journey and help provide a roadmap for progress across the maturity levels. The comprehensive process is described as follows:

**Data Awareness Review:** In this step, surveys and interviews are conducted across departments for the purpose of evaluating how well data is understood and its value properly captured and used in applications, valued, and used. The guiding rubric in this review process is whether employees are able to view data as a resource rather than just some bulk of information residing in their records and whether they feel empowered by the leadership teams to use it.

**Infrastructure Review:** In this step, the existing state of data systems are assessed, namely, the levels to which they follow state-of-the-art guidelines, whether they are sufficiently integrated, and whether they are scalable across departments. This step allows organizations to identify any potential remnants (legacy data systems) or data silos that are preventing democratic access or slowing innovation.

**Technical Capability Assessment:** In this step, a full auditing of the tools required for processing the data (including but not limited to, its analysis, visualization, and access channels) is performed. Through this audit, organizations can identify user needs: for example, whether they have the platforms required for their operations (e.g., BI tools, data lakes) and whether these tools are being used effectively.

**Governance Evaluation:** In this step, organizations conduct reviews for various governance issues beginning with existing policies around data access, data privacy, data management and oversight, as well as compliance. In

tandem with policies, this evaluation also examined whether roles (like data stewards, owners, or custodians of the data) are clearly defined.

**Access Framework Analysis:** In this step, organizations must perform a thorough investigation to understand how data access is granted and whether it is based on roles and responsibilities. The access framework analysis aims to eventually assess the ease by which users are finding the data they need to perform ensuing related operations.

**SWOT Analysis:** In this step, organizations use the preceding outcomes to holistically identify their strengths, weaknesses (e.g., outdated tools), opportunities (e.g., new training programs), and threats (e.g., compliance risks). Such areas could touch upon leadership issues, outdated tools, the need to mandate new training programs, or an abundance of data compliance risks.

The process described above helps align an organization's strategic priorities with data democratization goals. It will guide the set of recommendations we provide under the Data Democratization pillar in this working paper.

## Framework for Responsible AI

---

Responsible AI (RAI) is the practice of designing and using AI systems in ways that are ethical, transparent, fair, and aligned with human values and societal well-being (Akbarighatar et al., 2023; Reuel et al., 2024; Voenekey et al., 2022). It ensures that AI respects privacy, avoids bias and discrimination, supports accountability, and is used for socially beneficial purposes. Responsible AI also involves diverse stakeholders beyond the technical community, who develop and ensure regulations and ethical guidelines, and continuously monitor AI systems to prevent harm and build public trust.

This section draws on a comprehensive review of literature on responsible AI (RAI) maturity models and readiness frameworks, synthesizing findings from a systematic analysis of 1,451 publications (Akbarighatar et al., 2023). The sociotechnical perspective adopted here offers a theoretical foundation for translating RAI principles into actionable practices, aiming to balance humanistic and instrumental goals. The resulting maturity framework upon which we develop several key strategic recommendations presents a holistic approach to organizational RAI capabilities, while acknowledging that pursuing human-centered outcomes in AI development may present greater challenges than in other technological domains. Additionally, we account for the complexities introduced with the advent of generative AI and agentic AI, for which various governance imperatives must be observed, and around which substantial ambiguity still lingers, considering the increasingly large gap in the speed by which those technologies are evolving, and by which human regulators are able to account for all the risks and compliance matters.

## Ethical AI

---

Responsible AI and ethical AI are closely intertwined, each offering complementary principles for guiding AI development. Responsible AI emphasizes accountability, transparency, and regulatory compliance, while ethical AI focuses on fairness, privacy, and societal impact. Integrating both is essential and being mindful of how they complement each other is recommended: just as ethical intentions need responsible implementation, responsible AI practices must be rooted in ethical values. Together, they enable organizations to build AI systems that are legally compliant, value-driven, and designed to minimize harm.

### ***Accountability Risks***

The designers and deployers of AI systems need to be accountable for the operations of their systems, particularly when their decisions affect people's lives. There should be a clear and defined chain of accountability across various stages of the AI system's life cycle, from design and development to deployment and maintenance, ensuring that those responsible can be traced back to any decision that affects individuals. Error accountability is needed due to a variety of compounding factors, starting with the inherent unpredictability of probabilistic ML systems, to hallucinations inherent in GenAI applications, to jailbreaking risks where users intentionally bypass the model's safety and content restrictions in order to make it generate harmful output:

The accountability of AI systems should be designed to ensure that humans are in the loop on the decisions made by the AI models. Organizations must be equipped with proper regulations that clearly define who is to be held responsible if AI makes a harmful decision or recommendation, and how do we audit and explain AI-driven outcomes for regulatory or compliance purposes.

### ***Security and Privacy Risks***

Protecting individuals' privacy and personal data is essential for organizations. This involves ensuring that personal information is handled responsibly and safeguarded against threats. Privacy is concerned with how data is collected and accessed, while security focuses on protecting it from harm. Privacy becomes of paramount importance as organizations import existing GenAI solutions. To date, organizations are still grappling with the question of who owns the data used in generative models, and to what extent is it being used responsibly and securely? Proper localized regulations must address this schism and work to clarify the intricacies bearing impact on security and privacy in the age of GenAI.

Some unintended biases are increasingly emerging with the advent of GenAI applications and tend to be extremely nuanced. GenAI models can introduce subtle, hard-to-detect forms of bias. These may stem from training data, reinforcement signals, or model architecture. Bias can influence outputs in hiring, lending, healthcare, content moderation, and many crucial applications from the policy sector. This requires significantly distinct approaches for consumer data protection (e.g. user consent, anonymization, and fair use), training data governance (e.g. clarifying ownership, consent, and diversity of datasets used to train AI models), and intellectual autonomy of human workers (e.g. preventing over-reliance on AI and preserving human decision-making and creativity).

### ***Algorithmic Bias and Inclusivity Risks***

Biases in AI systems stem from multiple sources, which makes mitigating them extremely challenging. One major contributor to AI bias is tied to the data collection process itself, where historical inequalities or societal prejudices tend to be embedded in the training data. Human influence also plays a role, as subjective choices during data labeling for supervised machine learning algorithms or other algorithm design decisions can introduce further bias. Large language models, in particular, are trained on vast, unfiltered internet data, making them susceptible to subtle stereotypes. These biases have already been reported to have been manifested widely in real-world applications such as hiring tools, law enforcement, bail decisions, and healthcare systems, which exacerbates the need to address them using highly robust guidelines and recommendations.

### ***Legal and Regulatory Risks***

Oversight must address legal and regulatory risks tied to training data, not just intellectual property. A transparency pressure is mounting on organizations as the use of shared GenAI (e.g., GPT, Gemini) may require disclosure of

training data to regulators or the public. The copyright risk that arises as a result of that is tied to the fact that large language models use copyrighted data without explicit permission. The legal status surrounding the fair use of this data varies globally, and differing laws create unpredictable liability and operational risks. With the rise of digital twins ecosystems, employees' outputs may be reused to train AI "replicas", raising ethical and governance questions on ownership, longevity, and benefit sharing. Machine attribution governance in the age of GenAI poses even greater confusion that needs to be regulated. As GenAI is widely used to create written, visual, and video content, arbitrary and unregulated attribution decisions affect trust: Should audiences know what content was AI vs. human made? Even more complex regulatory dilemmas are transpiring in AI systems involving Multiple AI Agents. As Agentic AI acts can act autonomously by planning, deciding, and taking actions without constant human input, legal and regulatory processes now need to handle autonomy and control boundaries, specifically to determine how much freedom AI agents should have. Delegation of authority becomes a legal and ethical issue, as one now needs to navigate questions such as: "Can AI represent a company in negotiations, hiring, or contracts?", and if AI agents are acting as "quasi-employees", how do we settle questions surrounding liability, insurance, and audit trails?

### ***Sustainability Risks***

AI systems have a legacy of being extremely energy hungry, due to their high demand for computational resources. Complex models with a large number of parameters and sophisticated architectures consume substantial energy. The choice of model significantly affects this footprint, and end users are faced with an infamous dilemma, whether to trade the more energy-intensive, yet highly powerful deep learning, NLP, and generative AI models, for simpler, less performant alternatives that yet tend to be more environmentally sustainable. The energy footprint is not solely tied to AI models. Large data centers require intensive cooling, leading to high water consumption, and high-performance computing infrastructure including servers and GPUs pose environmental challenges, ranging from resource extraction and manufacturing emissions to the accumulation of electronic waste.

## **Trustworthy Machine Learning**

Whilst responsible AI focuses on the broader implications that encompass the ethical, legal, and societal implications of all AI systems, trustworthy machine learning (ML) provides the technical foundation needed to achieve the broader goals of responsible AI.

AI practitioners would all agree to this grim reality: more often than not, AI systems are prone to fail, or produce harm, whether intentionally or not. This is because real life scenarios which AI systems are meant to navigate, are highly complex, volatile, and uncertain. Traditional machine learning pipelines tend to be focused on accuracy and related metrics (precision and recall, F1 scores, and ROC curves, in the case of classification, MSE or similar metrics in the case of regression, Bleu scores in the case of NLP, etc.). Those metrics all assume a static ground truth against which AI systems are benchmarked "in the lab". When deployed across real life applications, there is a growing need to design AI systems that adhere to guardrails from *trustworthy machine learning*, in order to prevent significant failures that can at times, be very catastrophic.

Trustworthy ML can be seen as a subset of responsible AI (Varshney, 2022) and focuses more narrowly on ensuring that machine learning models are fair (free from harmful bias), transparent (understandable in how they make decisions), robust (able to handle unexpected inputs or distributional shifts), and secure (resistant to attacks or manipulation). Trustworthy ML also emphasizes the ability to account for uncertainty or probabilistic outcomes, especially in high-stakes applications like healthcare, finance, or criminal justice. Small data.

Trustworthy ML is still an emerging subcomponent of responsible AI. The fact that it incorporates intricate statistical subtleties is perhaps making it harder to incorporate it into off-the-shelf and commercial AI *as a service* solutions.

The following five facets of trustworthy ML will orient several of our recommendations under the responsible AI maturity framework as we proceed to demonstrate the critical implications of each of those components:

- Explainability and Interpretability
- Distributional Robustness
- Algorithmic Fairness
- Uncertainty Quantification
- Machine Learning for Small Data

### ***Explainability and Interpretability***

Many advanced machine learning models, specifically those incorporating deep learning models or support vector machines, are termed *black box models*: those are opaque models by which the end user has a limited understanding of their inner workings. AI systems built around black models, including GenAI models, need to be explained and interpreted, for users to develop some trust in their decisions and outcomes. Explainability is about helping people understand *why* a model made a decision, for example: “You were denied aid because your income appears to be high.” Interpretability refers to understanding *how* the model works, for example: knowing when an AI model gives more weight to smoking than to obesity when estimating insurance policy costs. Explainability and interpretability help users in debugging AI systems, auditing them, and examining whether these models have picked up biases or spurious correlations in the data. Ignoring those two aspects of a machine learning pipeline can lead to unfair, confusing, or dangerous decisions, especially in high stakes situations.

Many well-developed tools can be used nowadays to improve explainability, most prevalent of which nowadays are SHAP and LIME. To improve interpretability, practitioners resort to models that are simple and transparent (e.g. decision trees) and striking a balance between predictive accuracy versus interpretability is a common strategy in trustworthy machine learning pipelines.

### ***Distributional Robustness***

Distributional shifts happen when the data that a model encounters when deployed in real life follows a significantly different distribution from that it followed during the training and testing phase. Sources of shifts could be related to fluctuating environmental conditions, evolving societal behaviours, or policy interventions. From a practical standpoint, this can happen due to events like economic crises, new diseases, or communities moving around. This causes a significant deterioration of performance for AI models. For example, a model trained on data generated from the greater Beirut area may mispredict the same outcomes for refugees in rural areas, or health models built for adults may mispredict for children. Distributional adaptation, hitherto referred to as distributional robustness, encompasses a collection of mitigation strategies to render the same AI models reliable again once a shift has been detected, be that across time, populations, or domains. To achieve distributional robustness, a machine learning pipeline is designed with shift detection mechanisms whereby input data is monitored for any deviations from the expected distribution. In lay terms, for sudden changes in patterns, like new populations appearing or unusual economic behaviors. At the data level, including more diverse training data, using data augmentation, or simulating crisis scenarios, help reduce the likelihood of shifts occurring. At the model level, methods to improve model generalization include domain adaptation, adversarial training, or tuning models that focus on stable features that don’t change across spatio-temporal or other contexts.

### ***Algorithmic Fairness***

Algorithmic fairness (or rather, the lack of it), happens when an AI model delivers worse results for certain groups or individuals based on protected attributes. Obviously, lack of algorithmic fairness leads to discrimination and unfair decisions in AI automated systems where real-life decisions hinge on the outcome reached via an AI decision. Examples of such adverse repercussions include but are not limited to, Skewed resource allocation, reinforcement of inequalities, misguided policy decisions or high stakes interventions. Algorithmic (un)-fairness can result from data collection bias (sampling bias, measurement bias, selection bias, or aggregation bias), or model development bias (reward functions, feature selection), or evaluation bias (choosing metrics that favor the majority in a group). To detect unfairness, one must compare how well the model works across different groups, for whether one subpopulation is subjected to more errors or worse outcomes than other groups. To reduce instances of algorithmic bias, one can investigate the role of biased features with the help of machine learning explainability and interpretability, apply domain adaptation or transfer learning to adjust for context surrounding a specific population, and apply imbalanced learning evaluation metrics in favor of minority groups. When those steps are practiced routinely, algorithmic fairness ensures AI systems treat different groups (like men and women, or rich and poor) equally, and that training across certain privileged subpopulations (e.g., wealthy areas in a city) is not to be generalized across the rest of the population.

### ***Uncertainty Quantification***

The world around us is uncertain by nature, and treating outcomes and decisions of AI models deterministically fails to capture such natural nuances, leading to inaccurate representations of knowledge. Uncertainty around us can be categorized into two types: *epistemic uncertainty*, that results from our limitations in our knowledge about the world. For example, if an AI model predicts that a bank should approve a loan application to a given customer, but they fail to repay the loan, this type of uncertainty could have been a result of the fact that the individual had a bad credit record, except this information was not collected earlier. Hence, epistemic uncertainty can be eliminated by obtaining additional information to feed to the predictive model. AI model uncertainty is also considered part of epistemic uncertainty. In contrast, *aleatoric uncertainty* refers to intrinsic uncertainty that cannot be avoided. In the example above, maybe the customer got robbed. Gathering additional information in this case would not have helped.

In light of the above, it becomes crucial for machine learning systems that are deployed in the real world, to reveal the extent to which they are uncertain, a process known as *uncertainty quantification*, so that a human in the loop can mitigate the circumstances when uncertainty is high. Many approaches for uncertainty quantification exist in the literature, but we choose to propose some of the latest approaches that will inform our recommendations within this working paper.

### ***Conformal Prediction Uncertainty Quantification***

Conformal prediction captures how certain a model is of its output by producing a given range of values, in the case of regression, or a confidence score, in the case of classification. For example, instead of predicting that a certain family requires exactly \$80 in aid, conformal prediction will indicate “between \$60 and \$100”; in the case of classification, it might indicate an output that says that it is “90% sure this household is high priority.” Without this type of predictions, machine learning pipelines risk over-trusting wrong predictions, as in giving too little aid in emergency situations. Similar metrics exist for Natural Language Processing models which are beyond the scope of this paper. For a detailed discussion on how to produce regression, classification, time series forecasting, and NLP models accounting for conformal prediction, we refer the reader to (Manokhin, 2023).

### ***Probabilistic Forecasting***

Managing risk involves making decisions that account for uncertainty and potential consequences, whereas aiming for accuracy focuses narrowly on predicting the most likely outcome. In extremely high-stakes fields, managing risk is often more critical than achieving maximum predictive accuracy. Probabilistic forecasting plays a vital role in this context by offering a full distribution of possible outcomes rather than a single point estimate representing one single outcome that end users may misconstrue to be binding. This allows stakeholders to comprehend the full range of risks and develop contingencies accordingly. Ignoring probabilistic forecasts in favor of point predictions can lead to overconfidence and poor decision-making. For instance, when predicting the likelihood of a patient developing sepsis, a model that outputs a single probability (e.g., “30% chance”) lacks the nuance needed to assess urgency and allocate proper resources. A probabilistic forecast, on the other hand, might provide a distribution that reveals a significant tail risk, such as a 10% chance of rapid deterioration, for example, which could warrant closer monitoring or earlier intervention. In more ways than not, probabilistic forecasting is not just a technical improvement but a practical necessity for responsible, risk-aware medical decision-making. For a detailed discussion on how to produce predictive models that estimate conformal predictive distributions, we refer the reader to (Manokhin, 2023).

### ***Machine Learning for Small Data***

“Small data” refers to data that is limited in size, high-dimensional, expensive to collect, hard-to-collect, or rare to find. This is data at a human scale, that is intentionally collected, unlike big data, that comes in massive volumes and at a high rate. Small data is not only limited to data deserts but can also be found in data mature environments depending on the application (e.g. cancer prediction, earthquake prediction, financial collapses, etc.). Small data poses a considerable change for AI practitioners on more than one front. AI models find difficulty to learn meaningful associations or generalizing well using small data. There is a higher risk of distributional shifts and algorithmic bias.

Machine learning processes for small data can be data-centric (data augmentation or interpolation), model centric (using transfer learning from similar domains, or few shot learning algorithms), cost centric (using cost sensitive machine learning), or evaluation centric (using model evaluation metrics that give importance to the minority groups or rare instances in the data).

## **Framework for AI Value Creation, Capture, and Destruction**

The journey towards the path of data and AI maturity is a kind of self-reflection. It requires us to look inward and ask, how prepared we are to embrace AI. With each step forward, however, we want to ensure that we are able to *create* AI value. Here, AI does not just automate tasks. To create value means to help stakeholders make better decisions, uncover patterns, predict demand, and deliver smarter, faster services, to create new products, to improve efficiency, and to curb unforeseen losses.

Pushing this paradigm even further, we emphasise that innovation alone is not enough. In fact, we aspire that our national AI strategy enables an ecosystem to master AI *value capture*. This is when all the innovation and productivity translate to real economic and strategic advantage, where Lebanon begins to secure patents, protect its data and models, and use its AI-powered insights to outmaneuver competitors. This is where profits rise, market shares grow, and leadership in the AI industry becomes undisputed.

In the following, we adopt the analysis by (Åström et al., 2022) in identifying the prerequisites for AI value creation and value capture, and use their treatment to develop recommendations on how to design the value-capture mechanism described in this section.

### *AI Value Creation*

Before organizations can effectively create value with artificial intelligence, they must build an ecosystem that supports its use, through organizational, technical, and industry-specific conditions that enable it to function meaningfully. We explore four key areas, or "second-order activities", that must be addressed to build a strong AI value-creation network (Åström et al., 2022).

The first activity lies in assessing internal AI maturity. This means understanding how advanced an organization is in its use of AI, from its technical knowledge to its ability to apply AI tools effectively in the presence of its own internal limitations or uncertainties and fluctuations brought about by external factors.

The second activity revolves around evaluating infrastructure needs for acting on automated outputs. AI does not create value by merely producing predictions. Value is created when the organization takes timely and appropriate action based on those predictions. For example, if AI forecasts a network failure but no action is taken, nothing is gained. A consequence of this observation is that organizations must build the operational capacity to act on AI's output, be that through yet more automated processes within a cloud-based solution incorporating MLOps requirements, or with the help of a human in the loop.

The third activity involves considering industry and customer readiness. Different industries have different regulations, especially concerning data privacy, which can limit the mechanisms by which data is collected and used for AI models. Customers, also, vary in their willingness or ability to share data and act on AI-generated insights. No matter the accuracy surrounding a given AI system's performance, value can only be created if the customer trusts and acts on that information. This might require dedicated investments to calibrate data collection and usage mechanisms, change management for maximising benefits, or cultural shifts, factors that can slow down or prevent AI value creation.

The final activity is around conceptualizing value creation opportunities. Organizations must translate their technical understanding of AI into clear business benefits. These can come in three forms: cost savings (e.g., through automating repetitive tasks), increased revenue (e.g., through improving efficiency or product quality), and broader business gains (e.g., through better strategic decisions, increasing resilience in the face of disruptions, or reducing unexpected losses).

### *AI Value Capture*

To capture value effectively, organizations must build AI capabilities that align with industry needs and understand the sector in which they operate. Combining technical AI maturity with deep industry insight helps identify real customer needs and tailor solutions that are useful and impactful. However, this is only possible if customers are willing to share data and do not have unresolved privacy or security concerns. It thus becomes imperative that a strong data strategy that clearly explains how data will be used, stored, and protected, is a basic requirement to be able to capture AI value.

Second, organizations must design value delivery routines, in other words, how they actually get the AI solution to the customer. Two main delivery models are discussed in (Åström et al., 2022): outcome-based models and licensing models. In outcome-based models, providers work closely with customers on improving the AI system in alignment with the customer's environment. This close partnership builds loyalty and allows for deep customization but is harder

to scale since it requires high engagement from both sides. In contrast, licensing models offer ready-made AI tools that customers can buy and use with minimal involvement from the provider. This model has been seen to offer heightened scalability and profitability over time. Ideally, it is recommended that providers combine both models, engaging deeply with key clients for high-value outcomes while using licensing to reach broader markets.

One must bear in mind that both types of models are associated with different types of impending risks. In outcome-based models, a major risk is that customers may fail to provide sufficient data for a thorough model training or may disagree with how value is measured, thereby compromising the ability of and trust in the AI system. In licensing models, the control over how AI is used and what kind of performance it is delivering is delegated to the customer, which in turn, leads to a strategic risk: the simpler licensing model may “cannibalize” or reduce the appeal of the more complex, higher-value outcome-based services. Companies must therefore carefully balance both models to avoid harming their own businesses, which necessitates ongoing adjustments to AI capabilities, delivery methods, and pricing strategies.

### ***AI Value Destruction***

Artificial intelligence, while often seen as a powerful tool for solving global problems, can also unintentionally harm sustainable development. This concept is referred to as “AI value destruction,” and it occurs in two major ways. First, AI systems can fail to address the grand challenges they were designed to solve, such as climate change, social inequality, or poverty, due to poor design, misplaced priorities, or flawed implementation. Second, even when AI succeeds in tackling a specific issue, it may create new and unintended problems, such as reinforcing inequalities, misusing resources, or introducing harmful side effects during its development or deployment.

Mancuso et al., 2025, argue that the very qualities that make AI so powerful, such as its adaptability and generative capacity, also contribute to its potential for harm. Because AI systems can be used in many different ways and tailored to serve various goals, they often end up producing outcomes that are in tension with one another. For instance, a single AI system might serve the interests of some stakeholders while inadvertently disadvantaging others, or it might solve one part of a problem while deepening another. This creates a paradox: AI’s flexibility allows it to support sustainability goals, but that same flexibility can undermine them. In other words, AI can simultaneously create and destroy value.

The key to managing this paradox lies in how we understand and manage the grand challenges AI aims to address. Organizations that expand their understanding of these complex issues, by challenging outdated assumptions and rethinking their goals, are more likely to use AI in a way that avoids harm. However, if organizations rely on limited or rigid understandings of social, environmental, or economic problems, their AI innovations may amplify rather than reduce those problems. Most importantly, we note the increasing trend among experts in decision sciences and management sciences nowadays that contend that AI may not always be the best solution; in some cases, traditional, non-digital approaches might be more effective or more affordable.

Ultimately, we stress the importance of recognizing both the positive and negative potential of AI. The field has often focused too much on the promise of innovation without acknowledging the full impact it can have on sustainability. The suite of recommendations we provide under this pillar call for a more balanced and cautious approach, one that considers not just whether AI can be used to solve a problem, but whether it should, and whether less innovative solutions can do a better job.

## Framework for Human Labor in the Age of AI

According to emerging labor theory studies, the long-term success of a national AI strategy will depend not only on technological capability but also on adaptive and human-centered policy frameworks that promote the flourishing of human agency in the AI era (Ganuthula et al., 2024). As agency-driven work environments emerge, individuals are no longer just executing tasks; they are agents involved in designing and guiding AI systems whilst exercising human, rational judgment, labor policies must be reimagined. Rooted in Agency-Driven Labor Theory (ADLT), the transformation advocated by recent literature challenges traditional approaches to education and labor regulation. As AI increasingly handles repetitive functions and automated tasks, the remaining forms of human labor will demand high-level decision-making and strategic malleability. Policymakers must design responses that foster human agency while ensuring equitable access to opportunity in AI-enhanced economies.

Whilst the impact on existing educational systems, including higher education and professional and continuing education, are straightforward and clear (for example, by requiring that education systems be overhauled to meet the demands of AI-augmented work through strategic thinking, ethical reasoning, and the ability to design and adapt frameworks in uncertain, dynamic environments, perhaps the two other most affected areas Labor policy and Licensing systems.

### *Labor Policy*

Existing protections in labor policy have been previously designed for stable, well-defined roles; agency-driven work often crosses these boundaries, blending employment and entrepreneurship or straddling public-private spheres. Labor frameworks must evolve to address non-linear work arrangements, offering both flexibility and protection. This includes updating employment classifications, redefining benefits eligibility, and extending rights around intellectual property and ethical responsibility for system design.

Equitable value distribution is another critical challenge falling under Labor policy overhaul requirements. In agency-driven settings where individuals end up creating substantially high-impact systems, traditional compensation models such as hourly wages may fail to reflect true value creation harnessed by individuals. New mechanisms, such as platform equity, profit-sharing, or guaranteed minimum income, may be needed to ensure fairness across labor constituencies. Overlooking this very specific requirement might lead to economic rewards of AI being concentrated among a small set of highly empowered individuals, exacerbating already existing wealth and income inequalities in a country like Lebanon.

### *Licensing Systems*

Traditional credentialing and licensing systems are exclusively focused on the act of verifying technical knowledge or regulatory compliance. In an agency-driven economy, it becomes equally important to assess individuals' ability to design, lead, and act ethically in complex environments. New certification models that are based on peer review, demonstration of relevance to real-world projects, or the ability to maintain dynamic AI and data science portfolios, may better capture these emerging competencies.

Considering the above overhauls, one must require a forward-looking national AI strategy that has education reform at the heart to cultivate human agency, labor policies that address new work models, and certification systems that validate emerging competencies. A dedicated set of recommendations are put forth in this working paper, to account for the future of work implications.

## Final Recommendations

### Socio-materiality

1. Institutionalize intercultural dialogue, by embedding structured intercultural and pluralistic dialogue forums within AI policy development to surface local, indigenous, and marginalized perspectives.
2. Develop ethical frameworks grounded in pluralism and local values, where fairness is locally defined and informed by community-based consultation, whilst rejecting universalist approaches that perpetuate exclusion.
3. Maintain iterative, participatory governance structures that evolve as AI technologies and social responses change.
4. Strengthen political and affective communities, by supporting grassroots movements and advocacy groups working to democratize AI and contest technological hierarchies, as well as fostering “political friendship” and solidarity by creating spaces and funding initiatives that promote collaboration across divides.
5. Prioritize crowdsourcing as a strategic approach to AI development and deployment. By actively designing AI initiatives that incorporate crowdsourced data collection, validation, and problem-solving, the strategy can tap into the creativity and agency of citizens while addressing institutional and infrastructure gaps.
6. Embed Decolonial Principles in Governance, by explicitly incorporating decolonial goals, setting solidarity over paternalism, and inclusion over hierarchy. Incorporate contrapuntal analyses to identify and dismantle colonial binaries (e.g., metropole/periphery, West/rest, Rich/poor, or any other binaries at the level of subpopulations) in AI policies and systems.
7. Treat data not just as raw input but as a social artifact and regulate its collection and use with attention to power, consent, and cultural relevance.
8. Position crowdsourcing as an alternative to resource-intensive IoT systems. Given the high costs and maintenance demands of deploying sophisticated IoT networks, crowdsourcing offers a cost-effective, scalable alternative. Citizens can serve as “human sensors,” contributing valuable, timely, and context-specific data that may even surpass what rigid technological systems can provide. This approach also reduces dependency on imported hardware and proprietary systems.
9. Promote education and training that equips AI developers, policymakers, and citizens with the critical skills to recognize and address colonial continuities in technology.
10. Promote interdisciplinary education that includes social sciences, ethics, and local knowledge, alongside
11. Build capacity for managing and sustaining crowdsourced initiatives. Successful crowdsourcing requires thoughtful design of communication patterns, task distribution, and feedback loops. Training programs for public servants, NGOs, and local leaders should include principles of collective intelligence, ethical engagement, and the integration of AI tools to support such initiatives.
12. Design material infrastructures (e.g. cloud, energy, connectivity) with attention to local social impacts, especially environmental, economic, and labor issues.

13. Support bottom-up, community-led innovation rather than relying solely on top-down or foreign-led technological transfers. Encourage participatory action research and community-engaged design practices in AI development and promote the funding and institutionalizing of tools like citizens' juries and diverse voices panels to co-create AI policies and technologies.
14. Invest in tools and platforms to facilitate effective collective intelligence. While crowdsourcing reduces dependence on expensive infrastructure, it still requires accessible digital platforms, clear guidelines, and adequate incentives for participation. The national strategy should allocate resources to develop or adapt platforms that support collaborative work, equal participation, and transparent feedback mechanisms.
15. Mandate transparency and documentation, by requiring datasets and AI models to include clear documentation (e.g., data sheets, model cards) that explicitly state their assumptions, biases, and limitations, and by including mechanisms for affected communities to review and challenge these assumptions.

## Data Democratization

1. Establish a national Data Governance and Integrity Framework, representing the legal and institutional framework that defines data rights, AI ethics, and protections for whistleblowers who expose misuse, manipulation, or discrimination in AI systems. Enact whistleblower protection laws specific to data misuse and algorithmic harm and establish an independent data ethics board to review and audit high-risk AI applications.
2. Build modular, interoperable Data Infrastructure with crowdsourcing integration to supplement and accelerate national data efforts. Invest in open-source, cloud-based platforms that allow for community-based data contributions, especially in underrepresented sectors (e.g., agriculture, disaster response, public health). Embed SMS- or app-based data collection tools into government services and incentivize data contributions through recognition, microgrants, or community impact dashboards.
3. Establish a public Open Data and Citizen Science portal, by dedicating a national citizen science platform where individuals and institutions can contribute geotagged, observational, or sensor data relevant to national goals (e.g., community violence, corruption, pollution). Consider practical instruments for sensing by crowdsourcing such as low-cost toolkits (e.g. Arduino or mobile sensors).
4. Roll out national Data and AI literacy initiatives, highlighting the role of informed citizenry in understanding and contributing to data systems, whether through crowdsourcing workers or whistleblowers. Embed data literacy, ethical AI, and civic tech topics into national curricula and civil service training. Host national "Data for Good" challenges to engage youth and NGOs and launch community data clinics in rural and urban areas using mobile units.

## Responsible AI

### Ethical AI

#### *Accountability Risks*

1. Mandate clear chains of responsibility, by establishing legal and regulatory frameworks that require all AI systems to have a traceable chain of human accountability. Every phase beginning with design and training

to deployment and monitoring, should have named individuals or teams who are formally responsible for decisions, risks, and consequences.

2. Define liability for AI decisions, by introducing legislation that defines legal liability for AI-driven decisions that lead to harm or adverse outcomes. This includes both errors resulting from model unpredictability (e.g., probabilistic misclassifications) and misuse (e.g., model jailbreaking or hallucinated outputs). Responsibilities must be clearly assigned to AI developers, deployers, or operators, depending on the context.
3. Require human oversight and intervention mechanisms, and enforce policies ensuring that AI systems, particularly those affecting rights, welfare, or safety, operate under human supervision. Human-in-the-loop or human-on-the-loop mechanisms should be required for high-stakes applications in sectors such as healthcare, justice, education, and social services.
4. Standardize AI auditability and logging, by developing national standards for AI system auditability, including robust logging of model behavior, inputs, and outputs. These records must be preserved for post-hoc analysis, regulatory inspection, or legal redress. Mechanisms should be built into systems to explain decisions retrospectively and support investigations into model performance.
5. Establish independent AI oversight authorities, by creating a national AI oversight body empowered to investigate incidents, enforce compliance, and certify accountability mechanisms. This body should include technical, legal, and ethical experts capable of assessing complex decision chains in AI deployments.
6. Require impact assessments and risk disclosure by mandating regular AI impact assessments for systems that affect individuals or public resources. These assessments should include a review of accountability protocols, known risks (e.g., hallucination, jailbreaking), and mitigation strategies. Public and regulatory transparency of these documents should be required.

### *Security and Privacy Risks*

1. Enact comprehensive privacy regulations for GenAI systems, by developing specific legal standards that govern the collection, access, and use of personal data in GenAI applications, including mandatory consent, anonymization protocols, and limitations on secondary data usage. These regulations should clarify responsibilities for data protection across public and private sectors.
2. Define data ownership and use boundaries for generative models, by introducing legislation that explicitly defines who owns the data used to train generative AI systems, particularly when models are imported or hosted externally. Policies should require transparency around data sourcing and establish clear boundaries for responsible and secure use.
3. Mandate localized security standards for imported AI solutions, by requiring that all GenAI systems imported into national infrastructure, whether open-source or commercial, comply with localized cybersecurity, data sovereignty, and encryption standards to minimize vulnerabilities and ensure alignment with national security priorities.
4. Establish specialized governance for training data and bias mitigation, by creating a national body to oversee training data governance, responsible for evaluating dataset diversity, ensuring proper consent, and addressing the subtle forms of bias introduced through reinforcement signals or model architectures in GenAI tools.

5. Safeguard human autonomy in decision-making systems, by implementing regulations that preserve the intellectual autonomy of human workers by limiting the scope of AI-driven decisions in high-impact domains such as hiring, healthcare, and public policy. Establish guidelines for AI-human collaboration that prevent over-reliance on automated systems.
6. Strengthen cross-sector consumer data protection mechanisms, by developing a unified consumer data protection framework tailored for GenAI contexts, ensuring consistent practices around user consent, explainability, and redress across industries. This framework should support public trust by promoting fairness, transparency, and accountability in all AI-driven services.

#### *Algorithmic Bias and Inclusivity Risks*

1. Mandate bias audits for high-stakes AI systems, by requiring organizations to conduct and publicly report regular algorithmic bias audits for AI systems used in sensitive domains such as hiring, policing, healthcare, and finance, with metrics disaggregated by race, gender, socioeconomic status, and other protected attributes.
2. Establish inclusive data standards for AI training, by developing national guidelines for inclusive data collection practices that minimize historical bias and ensure adequate representation of minority and marginalized groups. This includes protocols for data diversification, annotation standards, and stakeholder input in data design.
3. Introduce certification requirements for fairness testing, by creating a certification system for fairness evaluation tools and methodologies. Any AI model deployed in the public sector or high-risk sectors must pass standardized fairness tests using approved frameworks before deployment.
4. Require human oversight in critical AI decisions, by enforcing the inclusion of human oversight mechanisms in AI systems making consequential decisions. Humans must have the authority to review, override, and explain AI outputs in cases involving individual rights or access to essential services.

#### *Legal and Regulatory Risks*

1. Mandate training data transparency standards by requiring organizations developing or deploying GenAI systems to disclose general categories and sources of training data, particularly in high-risk domains. Establish guidelines on how and when data transparency must be communicated to regulators, users, and affected individuals.
2. Clarify copyright and fair use in AI contexts by developing national legal guidance on the use of copyrighted content in AI training, including conditions for fair use and permitted exceptions. Introduce a registration or licensing framework for data sources to reduce legal ambiguity and cross-jurisdictional inconsistencies.
3. Define ownership and rights in digital twin systems by enacting legal safeguards clarifying ownership of employee-generated outputs when used to train AI systems that create digital replicas or functionally equivalent agents. Define rights to attribution, benefit sharing, and usage limitations, particularly in long-term deployments.
4. Establish attribution requirements for AI-generated content by introducing regulations requiring clear labeling of AI-generated written, visual, and audiovisual content, ensuring that end users can distinguish between human and machine-created work. Attribution guidelines should be standardized to support trust and accountability across sectors.

5. Regulate delegation of authority to AI Agents, by defining legal boundaries for agentic AI systems in decision-making roles such as hiring, contract negotiation, or public communication. Require human oversight checkpoints and clarify under what conditions an AI system may legally act on behalf of an organization.
6. Develop legal status frameworks for autonomous AI agents by launching a national consultation to assess the regulatory implications of agentic AI acting as “quasi-employees.” This should include proposals for AI-specific liability rules, insurance requirements, audit obligations, and legal personhood boundaries, adapted to varying levels of autonomy.

### *Sustainability*

1. Measure computational complexity by tallying operations or referencing model parameters and Large Language models or Agentic AI systems’ token usage, particularly for vendor-hosted models.
2. Estimate energy consumption based on the efficiency of the underlying hardware, using metrics typically provided by cloud service providers.
3. Quantify direct carbon emissions by referencing the carbon intensity of the data center or its Carbon Usage Effectiveness, while recognizing the difficulty of capturing indirect emissions.
4. Evaluate water consumption using on-site Water Usage Effectiveness and account for off-site impacts through Power Usage Effectiveness.

### **Trustworthy Machine Learning**

1. Consistently evaluate the feasibility of AI investments from a financial perspective, to make informed decisions and ensure that investments align with their business goals.
2. Build robust monitoring systems within MLOps for mitigating distributional shifts. Establish real-time data drift detection tools in all deployed AI systems. Monitor for shifts across populations (e.g., urban vs rural, nationals vs refugees) and domains (e.g., health, finance, education) and create early warning dashboards to alert ministries when models face unfamiliar data.
3. Diversify and future-proof national AI datasets against distributional shifts, by creating centralized, interoperable, and representative datasets covering all regions and vulnerable groups. Use data augmentation and simulation to model future shocks (e.g., mass migration, economic collapse), and prioritize data collection in low-data, high-risk zones, using mobile surveys or satellite proxies.
4. Ensure that AI models are trained Models for generalization and fairness. Promote domain adaptation and adversarial training in all public-sector ML models and enforce model evaluation on out-of-sample and underrepresented groups and subpopulations before deployment.
5. Adopt uncertainty-aware and risk-sensitive forecasting, by requiring probabilistic forecasts and conformal prediction intervals in high-stakes applications. Avoid binary decisions in crisis settings and ensure that AI systems express confidence in their class predictions whilst enforcing human review when models are highly uncertain.
6. Standardize explainability and interpretability across all AI systems, by mandating transparent model documentation and model explainability dashboards for expert review. Build citizen-facing explanation

systems to justify aid, benefits, or eligibility outcomes. Include appeal mechanisms backed by explainability tools, especially for marginalized populations, to contest outcomes of AI when bias or harm or even wrong decisions are suspected.

7. Address the adverse effects of small data and institutionalize imbalanced learning when needed. Support few-shot learning and Bayesian modeling for small, rural, or emergency datasets. Mandate imbalance-aware training and evaluation metrics (e.g., minority recall, subgroup performance, F2 measures or their equivalent). Incentivize local data partnerships with NGOs, municipalities, and CSOs to strengthen small data pipelines.
8. Institutionalize fairness audits and public oversight by creating an AI Ethics and fairness review board to audit government AI systems. Regularly publish bias reports, error disparities, and performance by group, and include civil society in model testing and algorithmic fairness assessments before scale-up.

## AI Value Creation, Capture, and Destruction

1. Guide the creation of inclusive value flow structures that clarify how AI-generated value, whether through cost savings, increased revenues, or broader business gains, is shared between providers, users, and the public. To maximize AI adoption, value structures must be flexible and responsive to sector-specific realities. This includes empowering local AI providers and their business development teams to identify win-win opportunities and design custom value propositions for various end-users, including small enterprises and community institutions.
2. Adapt formal contractual agreements to acknowledge the collaborative nature of AI value creation in a low-data context. Contracts must emphasize shared responsibility for data generation, preparation, and quality assurance. National strategy should offer model agreements that define roles for public institutions, private developers, and civil society organizations in the AI development cycle. These agreements must also recognize that data, AI algorithms, and their associated insights form a combined value proposition, and cannot be evaluated in isolation. To build trust and transparency in such arrangements, governments should develop standardized performance indicators that help measure both the value delivered by AI solutions and the responsibilities of each stakeholder in enabling that value.
3. Ground revenue models and pricing strategies in performance-based indicators rather than static service fees. Pricing mechanisms should link compensation to measurable outcomes, such as operational cost reductions, service delivery improvements, or user engagement increases. Where possible, government policy should incentivize providers to use value calculators or “proof of concept” dashboards that help public clients visualize and quantify the impact of AI. In contexts where the provider assumes greater risk (such as outcome-based contracts), cost-based pricing may be appropriate. In areas where AI consistently delivers measurable improvements, providers should be encouraged to adopt value-based pricing models that reward performance and efficiency rather than input cost. National innovation funds or public-private AI partnerships can help de-risk this process for early-stage projects.
4. Clearly define the major sustainability issue the AI innovation is intended to address, be that in climate change, healthcare access, education, inequality, or another area aligned with the UN Sustainable Development Goals (SDGs).

5. Once the challenge is defined, evaluate how the AI solution contributes to solving it. Ensure that AI solutions offer measurable benefits in environmental, social, or economic terms. Abide by quantifiable metrics whether the AI solution helps reduce emissions, increase inclusive access, improve equity, or enhance efficiency.
6. Examine risks of value destruction and critically assess the potential downsides. Investigate whether the AI system is failing to address the intended issue and whether it is inadvertently causing algorithmic bias, environmental overuse, or socioeconomic disruption.
7. Balance the expected benefits of the AI innovation against the potential harms. Consult a broad range of stakeholders, including developers, users, affected communities, and regulators, to gain a full picture of the innovation's likely effects.

## Human Labor in the Age of AI

1. Reform education systems to build human agency and framework design capabilities. National education policy must prioritize the integration of curriculum, pedagogy, and assessment methods that cultivate strategic thinking and the capacity for autonomous decision-making to be able to design and critically evaluate systems alongside AI. Public investment in faculty development, interdisciplinary learning models, and experiential education should support this transformation.
2. Redesign labor protections for the complex realities of AI-augmented work. Restructure labor policies to address the evolving nature of work in agency-driven environments, where roles are increasingly hybrid and decoupled from traditional employment classifications. Policy should provide safeguards and protection for all workers, especially those operating at the margins of standard job categories, while enabling the flexibility necessary for innovation and agency expression. This includes provisions for fair compensation, intellectual property recognition, and the ethical use of autonomous systems in collaborative settings (e.g. human-AI agentic systems).
3. Revise certification and licensing frameworks to reflect emerging competencies, by expanding credentialing systems beyond technical knowledge verification to recognize new forms of competence relevant to AI-augmented labor markets. Certification should include assessments of agency, ethical reasoning, systems leadership, and framework design expertise. Alternative forms of recognition, such as portfolios, simulations, and peer-reviewed projects, should be supported, reflecting the breadth of material outputs from AI-driven solutions.

## Addendum: Analytics for Corruption

1. Adopt a Complexity-Informed Approach to Corruption.  
Recognize corruption as a complex, adaptive system rather than isolated incidents.  
Develop analytics frameworks that capture its dynamic, networked nature, including feedback loops, emergent behaviors, and resilience, to design more effective interventions.
2. Build Predictive Models Based on Behavioral Patterns.  
Implement predictive analytics to identify public officials or actors with voting or decision-making patterns that align with those of previously convicted individuals. This can help prioritize investigations and increase accountability in political and legislative processes.

3. **Develop Corruption Risk Indicators in Procurement.**  
Use data analytics to develop objective corruption risk indicators for public procurement, especially in high-risk sectors such as defense.  
Tailor interventions based on whether risk is concentrated in central or peripheral actors, informed by procurement and contracting data.
4. **Analyze Tax and Financial Transaction Networks.**  
Deploy network science and machine learning to uncover irregular patterns in tax filings and financial transactions. Such approaches can help identify hidden tax evasion schemes and their links to broader economic crimes like money laundering.
5. **Map and Monitor Global and Local Financial Networks.**  
Apply tools that reveal suspicious activities within financial networks, focusing on small, tightly knit groups around intermediaries and offshore entities. This can uncover hidden flows of illicit money and the actors facilitating them.
6. **Create a National Corruption Analytics Hub.**  
Establish a centralized unit within the government dedicated to collecting, curating, and analyzing data related to corruption risks, behaviors, and networks.  
Equip this hub with expertise in AI, machine learning, network analysis, and social science to produce actionable insights.
7. **Leverage Case Studies for Local Context.**  
Adapt proven methodologies from international case studies to the Lebanese context, ensuring that models are sensitive to local institutional, cultural, and data realities.
8. **Promote Transparency Through Open Data and Public Dashboards, where appropriate.**  
Publish aggregated analytics findings to increase public awareness, deter misconduct, and foster trust.  
Visualization of corruption risk indicators and network maps can empower civil society and the media to hold actors accountable.
9. **Invest in Capacity-Building and Education.**  
Train public officials, auditors, and analysts in the use of advanced analytics tools and methods to detect, understand, and counteract corruption effectively. This should include technical training as well as an understanding of the ethical and legal dimensions.
10. **Integrate Analytics into Policymaking and Enforcement.**  
Ensure that insights from analytics feed directly into policy formulation, resource allocation, and enforcement strategies, making anti-corruption efforts more targeted, efficient, and evidence based.

## Conclusion and Final Remarks

For an indeterminate time to come, even as technological evolution related to artificial intelligence, together with other rapid scientific and technological transformations such as advances in biotechnology, quantum computing, nanotechnology, and the increasing interconnectedness brought by digital networks, accelerates and becomes more complex, two polarized visions dominate the cognitive and epistemological debate surrounding AI.

On one side, some start from the premise that the brain is not magical: it is a biological computer made of neurons, governed by physical laws. Therefore, in principle, it is simulable. If we map and model its structure (for example, through artificial neural networks), machines could eventually replicate or even surpass its functions. This hypothesis aligns with what is called the computational theory of mind: the human mind is an information-processing system that can, in principle, be simulated, or even improved upon, by artificial systems. Thinkers like Daniel Dennett argue that consciousness is not an ineffable mystery but a product of complex informational processes (Dennett, 1991). Under the right conditions, a sufficiently advanced AI could develop artificial consciousness. This position is rooted in functionalism in the philosophy of mind, advocated notably by Hilary Putnam and Jerry Fodor (Fodor, 1990; Putnam, 1960; Putnam, 1967), where artificial intelligence is viewed as an "emergent intelligence".

On the other side, philosophers such as Markus Gabriel reject physicalism and reductionism (Gabriel, 2017). Gabriel argues that the brain is not the mind, and the mind is not reducible to brain activity alone. Humans are not algorithms: he challenges the view that the mind is software running on the brain. In his work on New Realism (Gabriel, 2015a; Gabriel, 2015b), he defends the irreducibility of consciousness, normativity, and ethics to any material computation. True intelligence involves contextual, creative, and common-sense reasoning that machines cannot replicate. The human mind is characterized by meaning-making: it intends objects, concepts, and values. This intentionality - the mind's "aboutness" - cannot be generated solely by neurons or physical processes. In this sense, meaning and subjectivity transcend the physical substrate.

This debate, deeply rooted in the history of philosophy of mind, will not be settled anytime soon, and far from being futile, it is essential. It is important to remain engaged with its substance, its tensions, the multiplicity of arguments, and the perspectives it opens. Artificial intelligence is not merely a techno-scientific revolution; it is also an anthropological revolution, both global and uneven, that demands a continuous effort of meditation and reflection. This requires the dynamic participation of all sectors of research, production, and all those concerned with physical and moral well-being within any given society.

In shaping a national AI strategy, it is crucial to move beyond a "plug-and-play" mindset and the allure of mere proofs of concept. AI systems must be developed and deployed using highly sophisticated trustworthy guardrails around distributional robustness, uncertainty quantification, explainability and interpretability, and fairness, whilst AI ecosystems within data deserts must be deeply grounded in highly sophisticated subdisciplines of machine learning that are particularly tuned to small data, such as transfer learning, meta-learning, and imbalanced learning. AI systems must be developed and deployed in ways that are not only responsible and robust but also deeply entrenched within the societal context in which they operate. To gain public trust and legitimacy, they must embody decolonial principles, avoiding the reproduction of historical patterns of domination and dependency, and local legacies of corruption and lack of accountability. For AI to be truly intelligent, it must harness the power of collective intelligence, by amalgamating diverse human perspectives and knowledge with technological capabilities. And for AI to be genuinely democratic, our national strategy must ensure that people retain agency and control throughout the process. This could not have been more crucial as we caution against the dismissal of collective intelligence in a society that is still reeling from collective civil war trauma which remains to be arguably and debatably lingering in Lebanon.

To ensure that AI is not just an internal upgrade of operations, departments, and tools, that it becomes a competitive weapon, Lebanon must work to delineate the differences and requirements between data maturity and AI maturity, on one hand, and AI value creation and value capture, on the other. And in doing so, it must ensure that it doesn't inadvertently destroy non-AI value residing elsewhere, and that it overhauls its understanding and management of human labor in the age of AI.

This working paper is by no means an exclusive roadmap: it is rather meant to complement the official LEAP strategy launched by the state minister of Technology and AI in 2025 using elements from responsible AI and trustworthy machine learning. By adopting mutual shaping and socio-materiality as a conceptual framework for our working paper, we are inviting policymakers in Lebanon away from a technology-first future, towards an AI strategy that is also co-produced with society. Particularly, we advocate that any foreseen national AI strategies must reflect the relational dynamics between people, power, institutions, data, and infrastructure, paving the way for a suite of recommendations that promote AI development as a democratic, ethical, and situated process.

## References

- [Åström et al., 2022] Åström, J., Reim, W., & Parida, V. (2022). Value creation and value capture for AI business model innovation: A three-phase process framework. *Review of Managerial Science*, 16, 2111–2133.
- [Abu Salem et al., 2019] Abu Salem, F. K., Al Feel, R., Elbassuoni, S., Jaber, M., & Farah, M. (2019). FA-KES: A fake news dataset around the Syrian war. *Proceedings of the 13th AAAI International Conference on Web and Social Media (ICWSM)*, 573–582.
- [Abu Salem et al., 2019] Abu Salem, F. K., Jaber, M., Abdallah, C., Mehio, O., & Najem, S. (2019). A distributed spatiotemporal contingency analysis for the Lebanese power grid. *IEEE Transactions on Computational Social Systems*, 6(1), 162–175.
- [Abu Salem et al., 2019] Abu Salem, F. K., Khalil, A., Al Khatib Al Khalidi, M., Awad, S., Hamdar, Y., Sami, H., Diederich, J., El Hajj, W., & Elbassuoni, S. (2019, January). Mobile phone records for exploring spatio-temporal refugee mobility: Links with the Syrian war and socio-economic variations in Turkey. *Proceedings of the D4R (Data for Refugees) Closing Workshop*, Istanbul.
- [Abu Salem et al., 2021] Abu Salem, F. K., Al Feel, R., Elbassuoni, S., Ghannam, H., Jaber, M., & Farah, M. (2021). Meta-learning for fake news detection surrounding the Syrian war. *Patterns*, 2(11).
- [Abu Salem et al., 2021] Abu Salem, F. K., Chaaya, M., Ghannam, H., Al Feel, R. E., & El Asmar, K. (2021). Regression-based machine learning model for dementia diagnosis in a community setting [Conference presentation abstract]. *Annual Alzheimer's Association International Conference*, Amsterdam.
- [Abu Salem et al., 2022] Abu Salem, F. K., Jurdi, M., Alkadri, M., Hachem, F., & Dhaini, H. R. (2022). Feature selection approaches for predictive modelling of cadmium sources and pollution levels in water springs. *Environmental Science and Pollution Research*, 29(6), 8253–8268.
- [Abu Salem et al., 2024] Abu Salem, F. K., Awad, S., Hamdar, Y., Kharroubi, S., & Jaafar, H. (2024). Utility-based regression and meta-learning techniques for modeling actual ET: Comparison to (METRIC-EEFLUX) model. *Artificial Intelligence in Agriculture*, 14, 43–55.
- [Al Noaimi et al., 2021] Al Noaimi, G., Yunis, K., El Asmar, K., Abu Salem, F. K., Afif, C., Ghandour, L. A., Hamandi, A., & Dhaini, H. (2021). Prenatal exposure to criteria air pollutants and associations with congenital anomalies: A Lebanese national study. *Environmental Pollution*, 281, 117022.
- [Akbarighatar et al., 2023] Akbarighatar, P., Pappas, I., & Vassilakopoulou, P. (2023). A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. *International Journal of Information Management Data Insights*, 3(2), Article 100193.
- [Bachelard, 1938] Bachelard, G. (1938). *La formation de l'esprit scientifique : Contribution à une psychanalyse de la connaissance objective*. Éditions Vrin.
- [Bonney, 1996] Bonney, R. (1996). Citizen science: A lab tradition. *Living Bird*, 15 (4), 7–15.

- [Cui and Yasseri, 2024] Cui, and Yasseri, T. (2024). *AI-enhanced collective intelligence* (arXiv preprint arXiv:2403.10433v2). arXiv.
- [Denette, 1991] Dennett, D. C. (1991). *Consciousness explained*. Little, Brown and Co.
- [Doughman et al., 2020] Doughman, J., Abu Salem, F. K., & Elbassuoni, S. (2020). Time-aware word embeddings for three Lebanese news archives. *Proceedings of LREC 2020*, 4717–4725.
- [Džanko et al., 2024] Džanko, E., Kozina, K., Cero, L., Marijić, A., & Horvat, M. (2024). *Rethinking Data Democratization: Holistic Approaches Versus Universal Frameworks*. *Electronics*, 13(21), Article 4170.
- [Fattah et al., 2021] Fattah, Y., Nourallah, M., Wahab, L., Abu Salem, F. K., & Elbassuoni, S. (2021). An RDF data management system for conflict casualties. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, 4711–4715.
- [Fodor, 1990] Fodor, J. A. (1990). Fodor’s guide to mental representation. In J. A. Fodor, *A theory of content and other essays* (pp. 3–29). Cambridge, MA: MIT Press.
- [Gabriel, 2015a] Gabriel, M. (2015). *Fields of Sense: A New Realist Ontology*. Edinburgh: Edinburgh University Press.
- [Gabriel, 2015b] Gabriel, M. (2015). Neutral Realism. *The Monist*, 98(2), 181–196.
- [Gabriel, 2017] Gabriel, M. (2017). *I Am Not a Brain: Philosophy of Mind for the Twenty-First Century*. Polity Press.
- [Ganuthula et al., 2024] Ganuthula, V. R. R. Agency-Driven Labor Theory: A Framework for Understanding Human Work in the AI Age. *SSRN Preprint*, December 29, 2024.
- [Granados and Nicolás-Carlock, 2021] Granados, O. M., & Nicolás-Carlock, J. R. (Eds.). (2021). *Corruption networks: Concepts and applications*. Springer.
- [Haklay et al., 2021] Haklay, M., Fraisl, D., Greshake Tzovaras, B., Hecker, S., Gold, M., Hager, G., ... Vohland, K. (2021). *Contours of citizen science: A vignette study*. *Royal Society Open Science*, 8(8), Article 202108.
- [Halwani et al., 2019] Halwani, D., Jurdi, M., Abu Salem, F. K., Jaffa, M. A., Amacha, N., Habib, R. R., & Dhaini, H. R. (2019). Cadmium health risk assessment and anthropogenic sources of pollution in Mount-Lebanon springs. *Exposure and Health*, 1–16.
- [Irwin et al., 1994] Irwin, A., Georg, S., & Vergragt, P. (1994). The social management of environmental change. *Futures*, 26(3), 323–334.
- [Irwin, 1995] Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. London, UK: Routledge.
- [Irwin, 2015] Irwin, A. (2015). Citizen science and scientific citizenship: Same words, different meanings? In *Science communication today: Current strategies and means of action* (pp. 29–38). Nancy, France: Presses Universitaires de Nancy.

[**Miller, 2022**] Miller, K. (2022, March 21). The movement to decolonize AI: Centering dignity over dependency. *Stanford Institute for Human-Centered Artificial Intelligence*. <https://hai.stanford.edu/news/movement-decolonize-ai-centering-dignity-over-dependency>.

[**Labadie et al., 2020**] Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020). FAIR enough? Enhancing the usage of enterprise data with data catalogs. In *Proceedings of the 2020 IEEE 22nd Conference on Business Informatics (CBI)* (pp. 201–210). IEEE.

[**Mancuso et al., 2025**] Mancuso, I.; Messeni Petruzzelli, A.; Panniello, U.; Vaia, G. *The bright and dark sides of AI innovation for sustainable development: Understanding the paradoxical tension between value creation and value destruction*. *Technovation 2025*, 143, Article 103232.

[**Manokhin, 2023**] Manokhin, V. (2023, December 20). *Practical Guide to Applied Conformal Prediction in Python: Learn and apply the best uncertainty frameworks to your industry applications* (1st ed.). Packt Publishing.

[**Mohamed et al., 2020**] Mohamed, S., Png, M.-T., & Isaac, W. (2020). *Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence*. *Philosophy & Technology*, 33(4), 659–684.

[**Mourad et al., 2025**] A. Mourad, F.K. Abu Salem, and S. Elbassuoni > ``Detecting Gender Bias in Arabic Text through Word Embeddings'', in *PLOS One*, 2025.

[**Putnam, 1960**] Putnam, H. (1960). *Minds and machines* (machine-state functionalism). In H.-N. Castaneda (Ed.), *Intentionality, Minds, and Perception* (pp. 179–183). Wayne State University Press.

[**Putnam, 1967**] Putnam, H. (1967) *Psychological Predicates*. In Capitan, William H., Merrill, Daniel Davy, Art, mind, and religion, pp. 37--48: University of Pittsburgh Press.

[**Reuel et al., 2024**] Reuel, A., Connolly, P., Meimandi, K. J., Tewari, S., Wiatrak, J., Venkatesh, D., & Kochenderfer, M. J. (2024, October 13). *Responsible AI in the global context: Maturity model and survey* (arXiv:2410.09985) [Preprint].

[**Riedl and De Cremer, 2025**] Riedl, C., & De Cremer, D. (2025). AI for collective intelligence. *Collective Intelligence*, 4(2).

[**de Sherbinin et al., 2021**] de Sherbinin, A., Bowser, A., Chuang, T. R., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., Popescu, I., Shonowo, A., & Sivakumar, K. (2021, March 25). *The critical importance of citizen science data*. *Frontiers in Climate*, 3, Article 650760.

[**Simondon, 2005**] Simondon, G. *L'individuation à la lumière des notions de forme et d'information*; Jérôme Millon: Grenoble, 2005; originally published 1958.

[**Varshney, 2022**] Varshney, K. R. (2022, February 16). *Trustworthy machine learning* (1st ed.). Independently published. ISBN 979-8411903959.

[**Vega et al., 2023**] Vega, L., Mäkelä, M., & Seitamaa-Hakkarainen, P. (2023). Listening to the sociomaterial: When thinking through making extends beyond the individual. *Design Studies*, 88, 101203.

**[Voenekey et al., 2022]** Voenekey, S., Kellmeyer, P., Mueller, O., & Burgard, W. (Eds.). (2022). *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*. Cambridge University Press.

**[Zhang et al., 2023]** Zhang, B., Abu Salem, F. K., Hayes, M. J., Helm Smith, K., Tadesse, T., & Wardlow, B. D. (2023). Explainable machine learning for the prediction and assessment of complex drought impacts. *Science of the Total Environment (STOTEN)*, 38.

**[Zuboff, 2019]** Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

## Appendices

In the realm of significant constraints tied but not limited to scarce data, limited AI infrastructure, and entrenched inequalities, this section highlights sociotechnical successes that stand to attest that these challenges need not define the limits of what can be harnessed from AI. The following contextualized case studies involving the first author and their collaborators, illustrate how even with localized and modest datasets, it is possible to build sophisticated, impactful machine learning applications. These case studies constitute a deliberate indication that AI can be developed in ways that are decolonial, respecting local knowledge and priorities, and that leverage collective intelligence: drawing on the insights, creativity, and agency of various existing communities themselves, and using “small” and “messy” data that is largely from our region. Under *Research*, we present real life case studies using trustworthy Machine Learning tied to applications in healthcare, agriculture, content moderation, crisis management, environmental studies, and conflict studies. Under *Teaching*, we present two case studies tied to public policy analytics and corruption studies. Under *Outreach*, we describe an ongoing case study involving demystifying AI for good governance in the Arabic language.

### A. Case Studies in Research

In (Abu Salem et al., 2019), we present a data-driven and semi-automated fact checking and fake news detection mechanism, within the Syrian war. We first construct FA-KES, the first recorded fake news dataset around the Syrian war, generated from several media outlets representing mobilization press, loyalist press, and diverse print media, whose content is matched to information representing “ground truth” obtained from the Syrian Violations Documentation Center (VDC). In (Abu Salem et al., 2021), we pursue the automatic detection of articles from FA-KES using a suite of features that include a given article’s linguistic style, its level of subjectivity, sensationalism and sectarianism, the strength of its attribution, as well as its consistency with other news articles from the same “media camp”. The model-agnostic meta-learning algorithm (MAML) suitable for few-shot learning on not-so-big datasets achieves the best performance when benchmarked against baseline and vanilla machine learning models, attaining an accuracy of 89%. Our study is the first to consider the notion of consistency of articles with respect to ground truth, and to capture divisive ideologies by documenting the sectarian discourse, which tends to impact the credibility of news articles adversely.

In [Fattah et al., 2021], we develop CasualtIS, an RDF data management system that models conflict casualties’ data as an RDF graph and allows users to query such data using a SPARQL endpoint, with examples from the Syrian and Iraqi wars. CasualtIS also includes a template-based natural-language querying interface to support non-expert users and can be used for fact-checking certain claims about conflict casualties, aggregating casualties over time and location, and finding contextual information about casualties. In (Abu Salem et al., 2019), we explore mobile phone records from Turk Telekom against large scale events surrounding the Syrian war as well as socio-economic variations across Turkish provinces. We develop indicators for Syrian refugee mobility through the volume of phone calls, Shannon’s entropy, and the radius of gyration. Our analysis shows significant variance in mobility among refugees as opposed to citizens in cities with low socio-economic development, peaking at times associated with large scale events in the Syrian war, and within proximity to those events.

In (Abu Salem et al., 2019), we employ large scale graph analytics to perform a topological vulnerability analysis of the Lebanese power grid subject to random and cascading failures, using some generic representation of the grid handed over by the Lebanese ministry of Water and Power. The Apache Spark implementation maps the topology of the grid to a complex network and is built around a local structural characterization of the Lebanese power grid that reveals a level of decentralisation via numerous connected components. In (Halwani et al., 2019), we embark on a

data-driven investigation for health risk assessment for Cadmium (Cd) in water springs in Lebanon, and for potential sources of pollution like agricultural and industrial activities in the Mount-Lebanon governorate, a semi-urbanized area in Lebanon, that also suffers from a solid waste dumps crisis. In (Abu Salem et al., 2022), the suite of predictor variables is extended to include vehicular traffic intensity and precipitation indices around the springs, as well as the slope of each solid waste dump. Machine learning regression models developed on this feature space can accurately predict Cd levels, with features such as traffic and the slopes of the dumps surrounding water springs being of high importance, which in turn can provide a strong basis for a risk management strategy. In (Zhang et al., 2023), we propose an explainable ML pipeline using the XGBoost model and SHAP model based on a comprehensive database of drought impacts in the U.S. The interpretation of the models at the state scale indicates that the Standardized Precipitation Index (SPI) and Standardized Temperature Index (STI) contribute significantly to predicting multi-dimensional drought impacts. The patterns between the SPI variables and drought impacts indicated by the SHAP values reveal an expected relationship in which negative SPI values positively contribute to complex drought impacts. With heightened incidents of drought events worldwide, the constant need to increase agricultural production demands a more cautious assessment of irrigation needs, and thereby a more precise estimate of actual evapotranspiration (ETa). In (Abu Salem et al., 2024), we demonstrate that few-shot, meta-learning models (MAML) that are specifically designed for enhanced generalizability on not-so-big datasets, outperform basic machine learning models in upscaling ETa from two major in-situ towers, the Ameriflux and Euroflux. Using limited remotely sensed land surface data from the METRIC-EEFlux and limited climatic variables, we demonstrate that the chosen models can attain quantifiable utility within the utility-based-regression paradigm towards impactful practical considerations. Our initial explorations reveal that EEflux ETa deviates significantly from in-situ observations measured through the Ameriflux and EEflux towers ( $R^2=39\%$ ), in contrast to extremely accurate estimations produced by MAML, especially those that are rare in the target distribution, which has cost-benefit significance in relation to ETa approximation of very large ETa values that constitute a load on irrigation systems. Of independent interest, this study confirms that limited remotely sensed EEflux products contribute significantly to knowledge about ground truth ETa and can thus be of valuable use in settings where access to good quality and high-volume data is compromised.

In (Al Noaimi et al., 2021), we examine the association between maternal exposure to criteria air pollutants and birth defects (BD) risk and develop cost-sensitive, machine learning models customized for an imbalanced and skewed dataset from the Lebanese national birth registry, to help predict BD risk scores as a function of well-established predictors as well as exposure to ambient air pollution during prescribed windows of risk. Feature Selection analysis confirms that maternal exposure to PM2.5 and NO2 are important features for several birth defects recorded in the registry. Also, our predictive models can be robustly employed to help prioritize patients at risk because of high top k precision and top k recall and can be interpreted/explained using rules that have been endorsed by the research community on birth defects.

In (Abu Salem et al., 2021), we tackle a dementia diagnosis dataset generated from three community-based surveys conducted in Lebanon as part of a larger dementia cohort study. The ground truth dementia diagnosis was obtained using the 10/66 algorithm that relies on a battery of tests and several hundred features, achieving a sensitivity of 0.92, specificity of 0.951, and an AUC score of 0.97. We show that meta-learning approaches attain stronger performance using only twenty features, offering a more parsimonious solution that also beats cost-sensitive learning on the imbalanced dataset.

In (Doughman et al., 2020), we present a set of word embeddings learnt from three large Lebanese news archives, which collectively consist of 609,386 scanned newspaper images and spanning a total of 151 years, ranging from 1933 till 2011. The diversified ideological nature of the news archives alongside the temporal variability of the embeddings offers a rare glimpse into the variation of word representation across the left-right political spectrum. To train the word embeddings, Google's Tesseract 4.0 OCR engine was employed to transcribe the scanned news archives, and various archive-level as well as decade-level word embeddings were learnt via Word2Vec models. To evaluate the accuracy

of the learnt word embeddings, a benchmark of analogy tasks was used. In ongoing work, we exploit those word embeddings in a multitude of ways suited for content analytics. For generations, women have fought to achieve equal rights with those of men. Many historians and social scientists examined this uphill path with a focus on women's rights and economic status in the West. Other parts of the world, such as the Middle East, remain understudied, with a noticeable shortage in gender-based statistics in the economic arena. According to the socio-cognitive theory of critical discourse analysis, social behaviors and norms are reflected by language discourses, which motivates our study in (Mourad et al., 2025), where we examine gender-based biases in various occupations, as reflected through various textual corpora. Several works in literature have shown that word embedding models can learn biases from the textual data they are trained on, which can propagate societal prejudices that have been implicitly embedded in such text. To explore this further, we adapt WEAT and Direct Bias quantification tests for Arabic, to examine gender bias with respect to a wide set of occupations as reflected in various Arabic text datasets. These datasets include two Lebanese news archives, Arabic Wikipedia, and electronic newspapers in UAE, Egypt, and Morocco, thus providing different outlooks into female and male engagements in various professions. Our WEAT tests across all datasets indicate that words related to careers, science, and intellectual pursuits are linked to men. In contrast, words related to family and art are associated with women across all datasets. The Direct Bias analysis shows a consistent female gender bias towards professions such as nurse, house cleaner, maid, secretary, and dancer. As the Moroccan News Articles Dataset (MNAD) showed, females were also associated with additional occupations such as researcher, doctor, and professor. Considering that the Arab world remains short on census data exploring gender-based disparities across various professions, our work provides evidence that such stereotypes persist till this day.

## B. Case Studies in Teaching

### *Public Policy Analytics*

A plethora of techniques can help transform data from the public sphere into actionable information that can help inform policy and improve delivery of government services. In the public sector, data analytics can enable precision policy, allowing governments and social sectors to implement targeted programs to address society's biggest problems at scale. Data analytics for public policy encompasses various techniques that help stakeholder move from reactive to proactive mode, adopting strategies that optimize on their resources. This course, developed and taught by the first author as part of the AUB graduate online diploma in AI and Data science, exposes students to applications that demonstrate social impact and data-driven decision making in the field of public policy. Using publicly available datasets and a mix of tools covering exploratory analysis, predictive analytics, spatial analytics, and NLP, this course walks students through real life, practical examples that demonstrate the effectiveness of those techniques in the public realm.

#### *Topics covered:*

1. How governments make data-driven and analytical decisions.
2. Principles of descriptive analysis, causal inference, and prediction, to inform responsible governance.
3. Linking two real-world international sanctions lists through deterministic and probabilistic record linkage.
4. Exploratory and time series data analysis to identify trends and expose critical flaws in data in applications from public policy.
5. Informing economic development targeting as well as structuring household consumption patterns using clustering techniques.
6. Linear and multiple regression forecasting analysis for predicting housing market prices.
7. Logistic regression for capturing trends in healthcare coverage in the United States and supporting a predictive targeting campaign to reach uninsured individuals.
8. Supervised classification techniques for predicting the extent of storm damage.

9. Introduction to corruption networks.
10. Assessing networked corruption risks in European defense procurement.
11. Conducting geo-processing to extract meaning from spatial data and to perform geospatial machine learning.
12. Sentiment analysis and topic modeling for the public sector.
13. Bias, fairness, transparency, and privacy issues in data science applications for the public sector.
14. How to design for impact, establish decision points, and shape data science products for public policy

### ***Corruption Studies***

Corruption is increasingly understood as a complex, adaptive social system that evolves and self-organizes, involving ever-larger networks of individuals whose interactions generate nonlinear, cascading effects. Unlike isolated acts in the past, modern corruption emerges from dynamic networks that adapt, spread, and undermine institutions and resources. These networks grow more resilient as their diversity and openness increase, making them harder to dismantle and more likely to survive. Understanding corruption through the lens of complexity science reveals it as a phenomenon shaped by randomness, feedback loops, and emergent behaviors that require equally sophisticated, multidisciplinary strategies to address, offering new pathways for policy interventions.

In this module which the first author taught as part of the public policy analytics course at the American University of Beirut, students were exposed, and thereby implemented, several case studies from (Granados and Nicolás-Carlock, 2021), all of which can be mapped to analogous corruption scenarios that exist in Lebanon.

In the case study *Predicting corruption convictions among Brazilian representatives through a voting-history based network*, Tiago Colliri and Liang Zhao examine nearly 30 years of voting data from Brazilian legislators. They focus on detecting the emergence of corrupt clusters within the legislative network by developing a predictive model that estimates a representative's likelihood of being convicted for corruption or financial crimes in the future, based exclusively on the similarity between their voting record and that of previously convicted politicians.

In the case study *Networked Corruption Risks in European Defense Procurement*, Agnes Czibik, Mihály Fazekas, Alfredo Hernandez Sanchez, and Johannes Wachs investigate defense procurement processes to create an objective indicator of corruption risk. Their findings show that the risk indicator is higher for military contracts compared to other types of contracts, and that corruption risk is significantly elevated at the periphery in some contexts, while in others it is higher at the center. This case study highlights the connection between corruption networks and economic crimes rooted in corruption.

In the case study *Identifying tax evasion in Mexico with tools from network science and machine learning*, Martin Zumaya and colleagues apply these methods to analyze data from more than 80 million taxpayers and nearly 7 billion monthly invoice aggregations. They construct temporal networks, where nodes represent taxpayers and directed links represent invoices within specific time periods, revealing that tax evaders exhibit interaction patterns distinct from the majority. This case study on tax evasion links to another economic crime: money laundering.

From another perspective, Oscar Granados and Andres Vargas explore the large-scale structure of global financial networks, examining specific features that signal suspicious activities such as tax fraud, corruption, and money laundering. They find that such activities tend to occur in small groups, emerging within communities of financial intermediaries, non-financial intermediaries, and offshore entities.

## C. Case Studies in Outreach

### *Content Creation in Arabic*

Urban Analytica at SAIL (UA) is a community-led knowledge producer dedicated to promoting a democratic and modern social order in the region. By leveraging qualitative and quantitative, evidence-based approaches, UA advocates and acts in accordance with civil society values. UA's interdisciplinary approach integrates insights and technologies from social and political sciences, data science, artificial intelligence, graph analytics, and data visualization. UA advocate for Arab societies where informed, evidence-based decisions drive societal progress, transparency, and accountability.

Currently, Urban Analytica (UA) at SAIL is being hosted by the SAIL for Change Center at AUB. In turn, UA hosts contributions from undergraduate or graduate students, or activists and community members at large, in relation to UA's content creation segment. Volunteers would choose a topic of interest, in application to corruption, crisis management, critical security, or any other priority area of relevance to our communities. UA would assist the volunteers with researching the latest trends in how AI or data science help disrupt ways that such application areas are addressed nowadays. Volunteers, under the guidance of UA, develop simplified scripts in Arabic, and work to produce blogs and accompanying videos that get disseminated amongst their surroundings.

Volunteers are recognized as collaborators on the work produced, get a chance to describe this in their CVs, as well as a chance to explore state of the art technologies for public policy and the social good.

UA is currently on LinkedIn, FB, IG, and X.



Issam Fares Institute for Public Policy & International Affairs  
American University of Beirut P.O. Box 11-0236 Riad El-Solh /  
Beirut 1107 2020 Lebanon



961-1-350000 ext. 4150



+961-1-737627



[ifi.comms@aub.edu.lb](mailto:ifi.comms@aub.edu.lb)



[www.aub.edu.lb/ifi/](http://www.aub.edu.lb/ifi/)



[Issam Fares Institute for Public Policy &  
International Affairs IFI](#)



[aub.ifi](#)



[@ifi\\_aub](#)



[@ifi\\_aub](#)



AMERICAN UNIVERSITY OF BEIRUT

ISSAM FARES INSTITUTE FOR PUBLIC  
POLICY & INTERNATIONAL AFFAIRS

معهد عصام فارس للسياسات العامة  
والشؤون الدولية