Theorizing the regulation of generative AI: lessons learned from Italy's ban on ChatGPT

Francesco Gualdi
London School of Economics and Political Science
f.gualdi@lse.ac.uk

Antonio Cordella
London School of Economics and Political Science
a.cordella@lse.ac.uk

Abstract

Existing literature has predominantly concentrated on the legal and ethical aspects of government initiatives to regulate AI, often relegating the technological dimension to the periphery. However, the emergence and widespread use of generative AI models present new challenges for public regulators. Generative AI operates on distinctive technological properties which require a comprehensive understanding by regulators prior to the enactment of pertinent legislation. This paper focuses on the recent case of the Italian ban on ChatGPT to illustrate the public regulators' failure in acknowledging the unique characteristics intrinsic to generative AI, culminating in a flawed regulatory endeavour. By drawing on the findings of an exploratory case study, this paper contributes to the theoretical understanding of AI regulation, highlighting the discordance between the dynamism and fluidity of generative AI and the rigidity of regulatory frameworks. The paper contends that until this tension is effectively addressed, public regulatory interventions are likely to underachieve their intended objectives.

Keywords: generative AI, ChatGPT, regulation, law, ethics, data policy.

1. Introduction

The governance of Artificial Intelligence (AI) systems has become a relevant task for both governmental and public entities in recent years. Notwithstanding, with the advent of generative language models (such as ChatGPT), regulatory bodies are encountering an array of novel dilemmas and challenges due to the distinctive characteristics of generative AI, which deviate considerably from traditional AI. ChatGPT, which utilizes a transformer-based machine learning model, identifies patterns from a vast database of sentences. This process involves steps such as input embedding and positional encoding. Input embedding transforms each word in the input into a vector that artificially represents the input in the relationships with outer inputs (words). Positional encoding, on the other hand, adds information in the vector about the position of each word in the sequence. The model then uses a decoder to generate an output based on these encoded inputs. The decoder constructs a sequence of words (or parts of words) for the output sentence, referencing the patterns it recognizes from the input embedding and positional encoding processes. The relationships between the encoded inputs and the decoded outputs are quantified as vectors, which enable the model to navigate the complex, multi-dimensional relationships between different variables. This transformation from vector relationships to functional relationships allows the model to create human-like text based on the input it receives (Vaswani et al., 2017). It is crucial to note that while the model produces remarkably human-like text, it does not truly "understand" the text in the same way humans do. The model is highly proficient at identifying and applying patterns in the data on which it has been trained. The generative characteristic of ChatGPT is in its ability to produce outputs applying patterns in the data on which it has been trained from inputs provided by the users. As will be further discussed later in this paper, the specific generative attributes of ChatGPT and other AI-driven natural language processing technologies necessitate new regulatory approaches that are technologically savvy.

Existing literature on AI regulation has principally been centred around three dimensions of regulatory activity: legal, ethical, and technological (Jobin et al., 2019). However, a substantial gap has been observed in the extant body of knowledge. The technological aspect is significantly underrepresented and undertheorized relative to its legal and ethical counterparts (Cath, 2018). This research deficit pertaining to the technological dimension potentially omits vital facets of AI regulation, especially within the context of nascent generative AI systems.

URI: https://hdl.handle.net/10125/106631
978-0-9981331-7-1

This paper engages with the topic of public regulation of generative AI, underscoring the necessity for a distinct conceptualization of generative AI before enacting regulatory measures. The research is rooted in an exploratory case study that scrutinizes regulatory initiatives regarding ChatGPT in Italy. The case study unveils that a regulatory approach for generative AI, which exclusively contemplates the legal and ethical dimensions of the datasets and access without considering the procedures by which the dataset is constructed, encoded, and converted into vectorially weighted relationships, misses vital elements that hinder effective regulatory efforts. Such oversight necessitates a more nuanced and comprehensive perspective on the regulation of generative AI that includes understanding its underlying technological mechanisms. Given these findings, this paper aspires to make two discrete contributions to the field. Firstly, it intends to afford due consideration to the technological dimension of AI regulation, highlighting the unique technological attributes of generative AI. Secondly, the paper endeavours to provide a more nuanced theorization of AI regulation, one that captures the specificities inherent in generative AI. A comprehensive and precise theorization of AI is paramount to formulating and enabling regulations that can competently manage the complexities of generative AI.

The structure of the paper is as follows: section two critically examines the existing discourse on AI regulation and identifies the prevailing lacunae. Section three explicates the chosen research design. Section four introduces the case study on the Italian regulation of ChatGPT. Section five discusses the findings in the light of the technological specificity of ChatGPT algorithm. Section six consolidates the conclusions of the study.

2. Background

Over recent years, there has been an upsurge in the creation of bills, acts, and norms aimed at regulating AI (Chae, 2020). However, governments and international public organizations such as the European Union, despite considering AI regulation a political priority, still grapple with the nuances of addressing it (Erdélyi & Goldsmith, 2018; Veale & Zuiderveen Borgesius, 2021). Noteworthy uncertainties revolve around the instruments to use, the objectives of the regulation, and the logic to be applied. Moreover, the emergence and dominance of novel AI systems pose a significant challenge to regulations that must keep pace with technological evolution (Fenwick et al., 2016). Regulation can be defined as "(…) sustained and focused attempts to change the behaviour of others in order to address a collective problem or attain an identified end or ends, usually but not always through a combination of rules or norms and some means for their implementation and enforcement, which can be legal or non-legal" (Black & Murray, 2019). Although diverse regulatory strategies for AI are being explored, this paper primarily understands regulation as the implementation and enforcement of formal norms. A key characteristic of formal regulation is that it distinguishes what complies with norms from what does not (Mohr & Contini, 2011). The literature has progressively recognized the multifaceted challenges in AI regulation (Bareis & Katzenbach, 2022; Barocas & Selbst, 2016; Cath et al., 2018; Hacker, 2021; Scherer, 2016; Sun & Medaglia, 2019). Despite being fragmented across several domains, literature on AI regulation provides distinct contributions to the study of AI. Specifically, works on AI regulation draw upon three complementary "guiding forces": law, ethics, and technology, as pointed out by Floridi (2018). Scholars have documented the efforts of governments and international public organizations to adhere to these guiding forces in the law-making process that should result in AI regulatory frameworks (Cath, 2018). Figure 1 conceptualizes the extant literature on AI regulation.

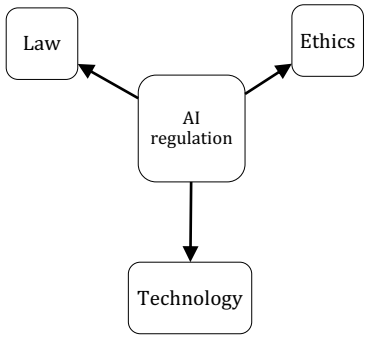

Figure 1. Dimensions of AI regulation.

The legal dimension of AI regulation focuses on the effort by regulators to design and implement legal frameworks governing AI (Black & Murray, 2019; Hacker et al., 2022; Surden, 2019). This debate emphasizes the different approaches taken by public organizations towards AI regulation. For instance, Cath et al. (2018) compare regulatory measures in the US, the EU, and the UK illuminating different governmental responses to AI regulation: the US with an antitrust approach, the UK adopting light-touch regulation, and the EU deploying a combination of soft and hard law with a risk-based approach (Floridi, 2022; Roberts, Cowls, Hine, et al., 2021). Moreover,

other examples of efforts to regulate AI emerged: for instance, China is adopting a tough stance against AI, requiring providers of AI services to perform security assessments and officially register their algorithms with the government (Roberts, Cowls, Morley, et al., 2021). Recently, attention has turned to the EU's efforts to regulate AI, given that European regulators have drafted the first legislation specifically addressing AI technologies (Hacker, 2021). Although still under discussion, the EU Artificial Intelligence Act (AIA) represents a novelty in the global regulations landscape (Floridi, 2022). It builds upon a risk-based approach that classifies AI systems according to the challenges AI use poses to society. The AIA shifts the regulatory paradigm as it is not technologically neutral as for example the EU's General Data Protection Regulation (GDPR) (Kesa & Kerikmäe, 2020).

The ethical dimension is of equal relevance to those studies investigating AI regulation (Tsamados et al., 2021). Ethical challenges arising in the regulation process focus on the significance of specific values such as fairness, transparency (Bringas Colmenarejo et al., 2022), moral responsibility (Tigard, 2021), privacy rights (Veale et al., 2018), accountability of AI designers and users (Gualdi & Cordella, 2021). Ethical concerns in this stem of research focus on the datasets that feed the algorithms (Floridi & Taddeo, 2016). Scholars focusing on ethical dimension of AI regulation urge regulators to adopt ethical frameworks when legislate on AI (Floridi, 2018). A good example of AI ethical framework is de Almeida et al. (2021) work that offers a comprehensive framework including as many different actors (policymakers, public regulators, business, judiciary, users) as possible in the call for regulating an ethical AI. Against this background, scholars have also raised warnings about the overreliance on ethical approaches towards regulation. For instance, Black and Murry (2019) posit that focusing on ethics when crafting a regulatory system is "essential, but by no means sufficient" (Black & Murray, 2019, p. 11) because it overlooks organizational and technological dimensions of AI adoptions.

While the legal and ethical dimensions have been thoroughly explored in the AI regulation debate, research placing the technological dimension at the forefront is relatively sparse. Moreover, as noted by Cath (2018), most research has often sought technical solutions to enhance AI regulation, such as algorithmic impact assessment (Reisman et al., 2018). These contributions, albeit important, overlook the significance of understanding the technology subject to public regulation. In line with Black and Murray (2019), AI regulators must engage deeply in understanding the construction of AI technology – the decisions made, the criteria selected to structure the technology (Black and Murray, 2019, p. 12). Given the rapid evolution of AI systems, the efforts to regulate AI might prove futile without an appropriate grasp of the myriad AI systems in existence (Helberger & Diakopoulos, 2023). The nature of AI as a technology is undertheorized in the debate about AI regulation, exposing regulators to potential risks and distortions in relation to the rapidly evolving nature of AI. This paper aims to address this gap, underscoring the relevance of the technological dimension in generative AI regulation and postulating that a proper theorization of AI is crucial to facilitating effective regulation. Governments and public organizations that overlook this dimension risk missing out on opportunities when designing and implementing key legislative pieces focusing on AI.

3. Methodology

The aim of this paper is to advocate for a new conceptual framework for understanding the burgeoning field of generative AI regulation. For this purpose, we utilize an exploratory case study approach (Yin, 2018) to illuminate the processes through which a public regulator enacted key measures to address ChatGPT. Exploratory research is a particularly suitable methodology when the research lacks a predetermined framework (Baxter & Jack, 2012), and the phenomenon under scrutiny is nascent or under-researched (Stebbins, 2001). Further, as Eisenhardt and Graebner (2007) note, an exploratory case study is more appropriate for novel theorization than are explanatory or descriptive case studies, which are better suited for theory-testing purposes.

The case study selected for this paper is the Italian regulation of ChatGPT. To the best of the author's knowledge, Italy is the first country in the world where a public body has imposed a normative measure on ChatGPT. While other countries (the US, the UK) or supranational organizations (the EU) are still considering, drafting, or discussing novel and specific regulation of ChatGPT (Hacker, 2023), Italy has relied on already existing regulation to assess the legal compliance of ChatGPT. The Italian case is specifically relevant because it has inspired other countries, with Canada being a remarkable example (Vermes, 2023). Moreover, the Italian regulation of ChatGPT constitutes a fitting example for an exploratory case study. Firstly, the events under consideration occurred in March and April 2023, and the case boundaries are still in flux with no definitive outcomes yet. Secondly, the case provides empirical evidence that can be used to "create theoretical

constructs, propositions, and/or midrange theory" (Eisenhardt & Graebner, 2007, p. 25). According to Yin's approach to exploratory research, data collection on a small scale should precede the formulation of the research question (Yin, 2018). Data collection encompasses secondary sources: (a) all the executive orders and media releases issued by the Italian regulator (8 documents); (b) General Data Protection Regulation (GDPR) bills (5 documents); (c) media releases (7 documents); and (d) documents issued by OpenAI (2 documents).

4. Case study

ChatGPT is a conversational AI model developed by OpenAI. This model specifically utilizes an extensive language database to generate contextually relevant text responses to inputted text prompts. Since its initial release in November 2022, this generative AI system has become a focal point in academic discourse and media discussions. Furthermore, its potential applications and associated risks have drawn considerable attention from regulatory bodies. Notably, the Italian Data Protection Authority, officially known as "Garante per la protezione dei dati personali" (henceforth referred to as "the Garante"), has been the most proactive in this arena. On 30th March 2023, the Garante – whose jurisdiction is limited to Italian territory – implemented a ban on the use of ChatGPT in Italy. The Garante's primary concern was centred around the compliance of ChatGPT with the European General Data Protection Regulation (GDPR) in terms of privacy protections.

4.1 Regulating ChatGPT

In Italy, a public independent regulatory institution is responsible for evaluating the conformity of citizens' personal data processing with legal mandates. The Garante is legally endowed with the capacity to inflict immediate punitive measures on entities that disregard data privacy norms. On 30th March 2023, drawing upon the principles and regulations established by the GDPR, the Garante issued an administrative decree imposing a provisional restriction on ChatGPT, barring the processing of personal data belonging to Italian citizens. The Garante's directive compelled OpenAI to instantaneously implement a geoblock on ChatGPT's operations within Italy. This course of action by OpenAI was the sole mean for the company to adhere to the Garante's mandate and circumvent potential legal prosecution. From the 30th of March until the 11th of April, ChatGPT was inaccessible for users located in Italy. The Garante found that OpenAI was in breach of GDPR principles. Specifically, the Garante found that ChatGPT was operating in violation of articles 5, 6, 8, 13, and 25 of the GDPR regulation. It is possible to cluster these violations under four main issues.

4.1.1. No information provided to users. The Garante articulates apprehension regarding OpenAI's disclosure practices pertaining to ChatGPT data collection practices. This is anchored in Article 13 of the GDPR, which stipulates that users must be informed when their personal data is being gathered. Specifically, OpenAI fails to clarify on key elements such as: the identity of the data collector, the purpose of data collection and processing, the nature of data sharing, and the legitimate interests being pursued. Users who utilize the ChatGPT service are not in the know to any of the aforementioned information. Furthermore, OpenAI has not issued any publicly accessible documents to elucidate the data collection methods used to feed and train its algorithms.

4.1.2. Accuracy of data processing. The Garante, drawing upon Article 5 of the GDPR, questions the precision of information disseminated by ChatGPT. Article 5 expressly exhorts private entities to maintain the accuracy of personal data and undertake substantive measures to purge inaccuracies. The Garante contends that the information generated by ChatGPT "does not invariably align with the truth". In the perspective of the Italian regulator, the erroneous or misleading information furnished by ChatGPT constitutes an infringement of Article 5, given that OpenAI has not demonstrated consistency in upholding the principle of accuracy during personal data processing. Furthermore, as highlighted by Hacker et al. (2023), the absence of accuracy could potentially give rise to discriminatory outcomes that contravene not only the principle of accuracy but also the principle of fairness. Therefore, in the view of the Garante, ChatGPT's alleged discrepancies underscore a critical need for stringent adherence to both accuracy and fairness principles, as mandated by the GDPR.

4.1.3. Absence of suitable legal basis for collecting users' personal data. The executive order put forth by the Garante unequivocally stated that OpenAI was functioning without "an appropriate legal basis" concerning data collection and processing. Particularly, the Garante interrogated the use of personal data to train the algorithms that constitute the foundation of the services offered via ChatGPT. This concern stems from a legal perspective anchored in article 6 of the GDPR, which requires a legitimate motivation for the collection and utilization of

personal data. The Garante's contention posits that OpenAI has not adequately provided or substantiated such a legitimate basis for their data practices in the context of ChatGPT.

4.1.4. Lack of age controls. The initial concern posited by the Garante pertains to the lack of mechanisms in place to ascertain that ChatGPT users are not below the age of 13, a criterion stipulated in OpenAI's own policies. There was, however, no protective system established for minors who engaged with ChatGPT. According to the Italian regulator, this omission contravenes Article 8 of the GDPR, which obliges private entities processing personal data to exert "reasonable efforts" to verify that consent is appropriately authorized by adults bearing parental responsibility. The Garante determined that the absence of age filters or specific tools could leave minors exposed to potentially problematic responses generated by the chat interface. This issue amplifies the need for more rigorous adherence to age verification and protection practices in line with GDPR guidelines, particularly when dealing with technologies that interact with diverse user demographics. Following several negotiation sessions between the involved parties, the Garante issued a subsequent order delineating the actions OpenAI was mandated to implement for the ban to be rescinded. The company pledged to comply with these requisites through numerous modifications to its website. In relation to the concern about information provision, OpenAI consented to divulge supplementary details about its data collection and processing practices via the publication of a dedicated document. To address concern about accuracy, OpenAI introduced a form through which users can request to access, delete, or amend their data. This form was prominently signposted via a specific link presented to users prior to their registration. To comply with concerns on absence of legal basis, OpenAI has equipped users with the requisite tools to assert their rights to opt-out from the processing of their personal data for the training of ChatGPT's algorithms when such processing satisfies the company's legitimate interest. Lastly, in response to concern regarding the lack of age controls, OpenAI has developed an age verification tool, acting as a gatekeeper within the interface. Upon assessing OpenAI's efforts to adhere to GDPR regulations, the Garante deemed them satisfactory and proceeded to lift the ban on 11th April 2023.

4.2 ChatGPT algorithm

The Garante intervention did not consider any technological feature at the core of ChatGPT system. An analysis of these core technological features reveals to be beneficial to better understand the challenges that a technological neutral regulatory approach faces when regulating generative AI systems. We will therefore analyse the specific feature at the core of the functioning of ChatGPT. The algorithm underpinning ChatGPT is based on three phases: pre-training of data, fine-tuning, and response generation. Pre-training begins with collection of text from huge datasets. The text is broken into small units of words – an action called tokenization. The algorithm is then trained to correctly identify random tokens based on context and words correlation. The model is further refined through training that focuses on predicting relationships between parts of text. Fine-tuning involves the creation of specific datasets that match possible inputs and related responses. Through supervised learning, the algorithm is trained to produce appropriate outputs: in this phase, the role of humans is crucial as they fine-tune algorithmic responses (Radford et al., 2018). Response generation happens when users access ChatGPT interface. Users' inputs are broken into tokens and encoded by the model. The algorithm then predicts the most probable sequence of words given the result produced by the processing of the user's input. The response is a sentence where the sequence of words is generated based on the most probable word series the system identifies given the way in which the context and information gathered is structured by the encoding mechanism used to codify the word and sentences in the training dataset. This action is crucial because likeliness of next words is determined through a probability distribution in the datasets (Radford et al., 2018). ChatGPT is able to produce meaningful reply to users' text queries because it uses and algorithmic modelling logic (Breiman, 2001) to predict responses. Algorithm modelling logic builds on the assumption that, in the case of natural language, sentences are the result of complex, mysterious, and partially unknown human cognitive processes. Therefore, it is not possible to predict what relationship among the multiple variables at stake defines the output. Hence, ChatGPT cannot replicate what many other AI systems using a data modelling logic (Breiman, 2001) do to build a stochastic data model to predict what the output will be.

The algorithmic modelling logic shifts the attention from how to formalise relationships among variables to replicate the univocal why by which the output is produced in nature towards the identification of multiple, probable combinations of variables that will be good predictor of what the output is. Hence, the capacity of the algorithm used to make the most probable prediction becomes crucial to define the

multiple possible outputs the algorithm can generate given the same set of inputs (Vapnik, 1999). Moreover, algorithms designed using the algorithmic modelling logic will produce multiple outputs (i.e, the same query submitted multiple times to ChatGPT will produce different results every time) since they generate the output based on multiple possible combination of variables associated to the input. What indeed determines the generative capacity of these algorithms is the endogenous capacity of the algorithm to produce multiple, different responses to the same query, where each response reflects similar or identical probabilities to predict a possible output.

All algorithms that use bootstrapping or other random selection of inputs to single out relevant variables or randomly selected combinations of variables to identify their predictive capabilities have therefore generative capacity. These algorithm do not build on explicit programming and rules-based approaches to solve problems and handle datasets (Helberger & Diakopoulos, 2023). Generative AI draws upon huge volumes of data: in the case of ChatGPT in the form of text fetched from several sources available online including websites, articles, books, social media data, to release outputs based on probabilistic computations on how they can possibly combine to respond to a given input (Bender et al., 2021). The responses generated by the algorithm also become new inputs in the dataset use by the algorithm to process further requests and elaborate responses.

Generative AI algorithms have therefore two dynamic elements at its core functioning. One is the expansive dataset which changes over time as the algorithm processes data to produce results. The other is the set of variables and weights used by the algorithm to produce responses. However, the variables and weights are changed by the algorithm over time to reflect changes in the dataset. In the case of ChatGPT, when the dataset changes, the probabilistic distributions of tokenized units and sub-units of text changes. In generative AI, outputs are unpredictable because they depend on mutable probabilistic text distributions. Furthermore, the algorithms itself generates unpredictability: the text it produces based on probabilistic distributions is also used as dataset for further elaboration.

5. Discussion

The choice of the Italian Garante to apply the GDPR to govern privacy-related aspects of ChatGPT is based on the lack of AI specific legislation – the EU's AIA is not enforced yet –, and the established assumption that the GDPR is technological neutral (Demetzou, 2019). Drawing upon a considered technologically neutral legislation such as the GDPR, the Garante aimed to regulate ChatGPT by requiring OpenAI to making the datasets and collection of personal data used to train ChatGPT algorithms compliant with legal and ethical standards of GDPR, and to regulate the access to ChatGPT so that it is compliant with the GDPR regulation. A comprehensive analysis of the impact of GDPR principles on generative AI is therefore essential to fully comprehend the significance of considering the technological specificity when formulating regulations.

5.1. GDPR principles and ChatGPT

The GDPR builds on 7 key principles: lawfulness, fairness, transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; accountability. Theses principle mainly deal whit how personal data of individuals is collected, processed, stored, and shared. The principles can easily apply to contexts where the mechanisms by which personal data of individuals is collected, processed, stored, and shared can be controlled and monitored. In response to these demands, OpenAI has striven to meet the expected data protection standards by disclosing additional information and implementing tools to prevent any violation of GDPR norms within its datasets. Despite these efforts, scholars such as Hacker et al., (2023) have highlighted limitations in this data protection approach when it comes to governing generative AI such as ChatGPT.

However, these criticisms do not provide a definitive answer as to whether the GDPR regulatory framework is capable of effectively addressing privacy concerns associated with the use of generative AI. The operational mechanisms underpinning ChatGPT pose in fact considerable challenges to the effective implementation of GDPR principles. Inherent in the design of the algorithm is the ability to generate multiple, diverse responses to an identical inquiry. Each of these responses is the product of a probabilistic analysis. The probabilities calculated by the algorithm, in turn, mirror the distribution of data within the dataset – specifically, the structural arrangement of sentences. However, the dataset is inherently dynamic. Namely, every output generated by the algorithm – derived from existing inputs – subsequently becomes a new input within the database. Each instance of new output production, therefore, results in an updated input within the dataset. Such modifications to the dataset precipitate alterations in the existing probabilistic distribution, which the algorithm then harnesses to generate novel

responses. This cyclical process undergirds the algorithm's generative capabilities and its capacity to learn, iterating continuously in response to its own outputs. This perpetually evolving process introduces significant complexity into the task of adhering to GDPR principles.

5.1.1 ChatGPT and the principles of lawfulness, fairness, and transparency. The expansive and complex nature of generative algorithms makes determining and maintaining a lawful basis for data processing more difficult. For instance, when an AI system generates new data from the original inputs, it's not always clear whether the lawful basis for the initial data collection extends to these new data points. Under GDPR, fairness means that data should be processed in a way that does not disadvantage or harm the data subject. However, the outputs from a generative model, particularly when based on large and diverse datasets, may inadvertently produce biased or discriminatory results. It can be difficult to predict, identify, and rectify such outcomes, thus challenging the principle of fairness. GDPR requires that data subjects are informed about how and why their data are being processed. However, the complex nature of generative algorithms and their capability to continuously generate and reprocess data can make it challenging to provide a clear, comprehensible explanation to the data subjects. Additionally, the dynamic nature of these algorithms makes it even more difficult to provide meaningful transparency.

5.1.2. ChatGPT and the principle of purpose limitation. The generative properties of the ChatGPT algorithm introduce substantial challenges in guaranteeing that personal data are procured for specific, explicit, and legitimate objectives, and are not further processed in a manner inconsistent with those objectives. In essence, ensuring that personal data are processed by the algorithm for a distinct purpose, transparently articulating that purpose, and limiting data collection to only what is necessary to fulfil that purpose, becomes an arduous task. The algorithm handles data as tokens, dissecting and reconfiguring them using probabilistic distributions that intrinsically modify the data, their intended use, and the context for their collection. For instance, the data collated for algorithm training transform into inputs to generate new outputs, which in turn generate additional data. This recurrent process modifies the objectives for which data are assembled and processed, potentially resulting in a manner incongruent with the initial objectives. Such an iterative mechanism raises questions about the ability to clearly define and maintain the integrity of the purpose for which the data was initially collected, given the algorithm's continuous learning and generative capability.

5.1.3. ChatGPT and the principle of data minimisation. The data minimisation principle stipulates that personal data collected should be "adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed". This principle necessitates a degree of restraint in data collection and processing, advocating for the least amount of data to be used to accomplish a given task. However, generative algorithms operate by leveraging vast amounts of data to generate nuanced and contextually relevant responses. In fact, the efficacy and accuracy of such algorithms are often directly proportional to the volume and variety of data they are trained on. Moreover, in the generative model, the production of new data is an intrinsic part of the process. Every output created by the algorithm is used as a new input, leading to a constantly expanding dataset. This dynamism and fluidity, while being a strength from an AI perspective, contradict the concept of data minimisation principle.

5.1.4. ChatGPT and the principle of data accuracy. The GDPR's data accuracy principle mandates that personal data must be "accurate and, where necessary, kept up to date". In essence, organizations have an obligation to correct or delete inaccurate data without delay. Generative algorithms, whose reliance on probabilistic analysis and the utilization of expansive, mutable datasets can lead to the production of outputs that may be inconsistent, imprecise, or misleading. For instance, ChatGPT might generate information that inaccurately represents an individual's personal data due to the data's contextual reassembly or erroneous initial inputs. Furthermore, the process of validating and updating the accuracy of data within a generative model is not straightforward due to its dynamic and self-feeding nature. As the volume and complexity of data increase with the algorithm's usage, so does the challenge of ensuring data accuracy.

5.1.6. ChatGPT and the storage limitation principle. This principle requires that personal data should be "kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed". In essence, organizations should not store personal data for longer than needed and should implement time limits for deletion or periodic review. Generative algorithms, however, operate on a continuous cycle of data usage. They not only consume vast amounts of data for training and functioning but also generate new

data that gets reincorporated into the system. This continuous data generation, reabsorption, and transformation process can make it challenging to adhere to the storage limitation principle. The expansive, generative nature of such algorithms makes it difficult to determine and enforce specific time limits for data deletion or review. Furthermore, the generated data – which effectively becomes a new input for the system – may indirectly contain personal data or identifiers, further complicating the matter.

5.1.7. ChatGPT and the principle of integrity and confidentiality. This principle mandates that personal data must be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures". The continuous cycle of data usage of generative algorithms change data over time, which can result in destroying or damaging personal data.

5.1.8. ChatGPT and the principle of accountability. The algorithm makes the task of documenting and explaining its data processing activities challenging. The dynamism of generative algorithms makes it difficult to keep track of and document data processing activities consistently. Generative algorithms can produce outputs that impact individuals, sometimes in significant ways, with minimal human intervention. According to GDPR, individuals have the right not to be subject to a decision based solely on automated processing if it produces legal effects concerning them or similarly significantly affects them. Under the GDPR, organizations are not only responsible for complying with the principles of data protection but must also be able to demonstrate their compliance. This includes, among other things, documenting data processing activities, implementing, and reviewing data protection policies, and ensuring data processing aligns with the GDPR's principles.

5.2 Beyond GDPR regulation: the technological dimension

The Italian regulator adopted an approach that overlooked the distinctive nature of generative AI when imposing restrictions on ChatGPT. The legal instruments utilized by the Garante to govern OpenAI disregarded the dynamic quality of datasets and algorithms deployed in generative AI services. The mandates issued by the Italian regulator to lift the ban on OpenAI encompass adherence to a rigid legal standard. This standard does not – and arguably cannot – accommodate the mutable and probabilistic essence of generative AI datasets and algorithms. For instance, the Garante required OpenAI to facilitate the process for users to rectify or erase their data when it propagates false or misleading information. In an effort to avoid contravention of Article 5 of the GDPR, OpenAI adhered to this requirement. However, the retroactive correction or removal of incorrect data does not guarantee the prevention of further dissemination of erroneous, misleading, or deceptive information by the generative model. The datasets upon which the ChatGPT algorithms are trained and fine-tuned continue to evolve and intermix, implying that potential inaccuracies in outputs are an inherent possibility due to the probabilistic nature of the output generation. This is indicative of the regulator's failure to consider the inherent dynamism of generative AI when formulating regulatory measures for ChatGPT.

The overarching strategy of the Italian Garante was rooted in a rigid understanding of regulation that sought to clarify what was in compliance (or not) with GDPR legislation. Nevertheless, this static logic collided with the dynamic nature of the datasets and algorithms upon which ChatGPT bases its services. Furthermore, given the characteristic dynamism of generative AI's datasets and algorithms, regulators should have taken this into account when implementing the measures to regulate ChatGPT. Within this context, the case study highlights an inherent tension between the rigidity of regulatory mechanisms, on one hand, and the fluidity of generative AI's datasets and algorithms, on the other. Figure 2 elaborates on the AI regulation dimensions to account for the dynamic properties of the technologies which underpin regulative AI.

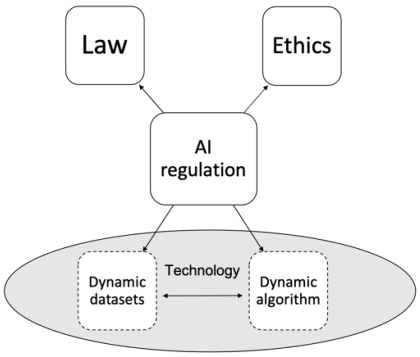

Figure 2. Framework that expands the technological dimension of generative AI regulation.

The case elucidates that regulatory frameworks such as the GDPR must adequately consider the technological underpinnings of generative AI to be effective. Indeed, generative AI, founded on

algorithmic modelling logic, challenges the conventional rationality that informs the design of digital technologies which have traditionally been the touchstone for the development of abstract legal frameworks applicable across various technological contexts. Regulatory frameworks, conceptualized with data modelling logic as a reference point, maintain technological agnosticism as long as the technologies they regulate are designed in line with the logic upon which these regulations operate.

The case discussed in this paper suggests that the shift towards algorithmic modelling logic, which drives generative AI technologies, poses substantial challenges to regulatory principles inspired by the regulatory necessities of technologies designed based on the data modelling logic. The implications of this paradigm shift necessitate a re-evaluation of our current regulatory frameworks, urging us to reconcile the dynamics of algorithmic modelling with existing principles of data governance and regulation.

This case study offers fresh insights to the Digital Government literature, highlighting the critical importance of integrating the technological dimension of AI. The case underscores that without a thorough examination of the technological underpinnings inherent to various AI systems, effectively addressing their impact on public sector progress and societal consequences, along with the outcomes of regulatory efforts aimed at managing such impact, becomes a complex task.

6. Conclusions

The distinct characteristics of generative AI models, such as ChatGPT, pose unique challenges for public regulatory authorities. The case study of the Italian ban to ChatGPT illustrates that regulatory intervention on generative AI with primary emphasis placed on legal and ethical concerns related to access to the system and to the structure of the dataset employed in training ChatGPT algorithms is flowed. The regulator articulated concerns regarding the absence of an appropriate legal foundation for gathering users' personal data and brought forward specific ethical dilemmas, encompassing potential discrimination, fairness, transparency, and the safeguarding of minors. Fundamentally, the regulatory strategy implemented by the Italian authority overlooked the unique technological facets of generative AI, which operates based on algorithmic modelling logic compared to conventional systems that operate on a data modelling logic.

Regulatory endeavours that neglect the inherent specificity of algorithmic modelling logic could potentially compromise the efficacy of the enforced measures. Considering the dynamic and mutable nature of perpetually evolving generative AI models based on algorithmic modelling logic, the rigidity of regulatory frameworks may prove inadequate for implementing measures. The Italian case offers relevant implications against the current landscape of increasing regulatory actions toward AI for those governments addressing data privacy in first instance (such as Canada) and for those public organizations aiming to produce a complete set of specific AI-focused norms (such as the EU) or AI-focused guidance (such as the US and the UK). The Italian case shows that in the absence of a comprehensive understanding and conceptualization of the underlying technology that drives generative AI, regulatory bodies may be unsuccessful in their attempts to deploy meaningful legislation for generative AI.

This research has attempted to shed a light in this direction, focusing on the first and most significant regulatory action against AI that was recorded. We are aware of limitations, such as the availability of data, that can impact our work. However, with this paper we aim to encourage the emergence of a debate in the Digital Government community, whose contribution is essential to capture the nuanced characteristics of AI, and to provide meaningful insights that can inform public policymakers' actions that regulate AI.

7. References

Bareis, J., & Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values, 47*(5), 855-881.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev., 104*, 671.

Baxter, P., & Jack, S. (2012). Qualitative case study methodology. *The qualitative report, 13*(4), 544-559.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?? Proceedings of the 2021 ACM conference on fairness, accountability, and transparency.

Black, J., & Murray, A. D. (2019). Regulating AI and machine learning: setting the regulatory agenda. *European journal of law and technology, 10*(3).

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199-231, 133.

Bringas Colmenarejo, A., Nannini, L., Rieger, A., Scott, K. M., Zhao, X., Patro, G. K., Kasneci, G., & Kinder-Kurlanda, K. (2022). Fairness in agreement with European values: An interdisciplinary perspective on ai regulation. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society,

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A:*

Mathematical, Physical and Engineering Sciences, 376(2133), 20180080.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': the US, EU, and UK approach. *Science and engineering ethics, 24*, 505-528.

Chae, Y. (2020). US AI regulation guide: legislative overview and practical considerations. *The Journal of Robotics, Artificial Intelligence & Law, 3*.

de Almeida, P. G. R., dos Santos, C. D., & Farias, J. S. (2021, 2021/09/01). Artificial Intelligence Regulation: a framework for governance. *Ethics and Information Technology, 23*(3), 505-525.

Demetzou, K. (2019). Data Protection Impact Assessment: A tool for accountability and the unclarified concept of 'high risk'in the General Data Protection Regulation. *Computer Law & Security Review, 35*(6), 105342.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of management journal, 50*(1), 25-32.

Erdélyi, O. J., & Goldsmith, J. (2018). Regulating artificial intelligence: Proposal for a global solution. Proceedings of the 2018 Conference on AI, Ethics, and Society,

Fenwick, M., Kaal, W. A., & Vermeulen, E. P. (2016). Regulation tomorrow: what happens when technology is faster than the law. *Am. U. Bus. L. Rev., 6*, 561.

Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180081.

Floridi, L. (2022). The European Legislation on AI: A brief analysis of its philosophical approach. In *The 2021 Yearbook of the Digital Ethics Lab* (pp. 1-8). Springer.

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*.

Gualdi, F., & Cordella, A. (2021). Artificial intelligence and decision-making: The question of accountability. Proceedings of the 54th Hawaii International Conference on System Sciences,

Hacker, P. (2021). A legal framework for AI training data—from first principles to the Artificial Intelligence Act. *Law, Innovation and Technology, 13*(2), 257-301.

Hacker, P., Engel, A., & Mauer, M. (2023). Regulating chatgpt and other large generative ai models. *arXiv preprint arXiv:2302.02337*.

Hacker, P., Naumann, F., Friedrich, T., Grundmann, S., Lehmann, A., & Zech, H. (2022). AI Compliance–Challenges of Bridging Data Science and Law. *ACM Journal of Data and Information Quality, 14*(3), 1-4.

Helberger, N., & Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review, 12*(1).

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389-399.

Kesa, A., & Kerikmäe, T. (2020). Artificial intelligence and the GDPR: Inevitable nemeses? *TalTech Journal of European Studies, 10*(3), 68-90.

Mohr, R., & Contini, F. (2011). Reassembling the Legal: 'The Wonders of Modern Science'in Court-Related Proceedings. *Griffith Law Review, 20*(4), 994-1019.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic Impact Assessments: A Practical Framework for Public Agency. *AI Now*.

Roberts, H., Cowls, J., Hine, E., Mazzi, F., Tsamados, A., Taddeo, M., & Floridi, L. (2021, 2021/11/12). Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US. *Science and engineering ethics, 27*(6), 68.

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & SOCIETY, 36*, 59-77.

Scherer, M. U. (2016). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law & Technology, 29*(2), 353-400.

Stebbins, R. A. (2001). *Exploratory research in the social sciences* (Vol. 48). Sage.

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly, 36*(2), 368-383.

Surden, H. (2019). AI and law: An overview. *Georgia State University Law Review, 35*, 19-22.

Tigard, D. W. (2021). Responsible AI and moral responsibility: a common appreciation. *AI and Ethics, 1*(2), 113-117.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: key problems and solutions. *Ethics, Governance, and Policies in AI*, 97-123.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.

Veale, M., Binns, R., & Edwards, L. (2018). Algorithms that remember: model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180083.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International, 22*(4), 97-112.

Vermes, J. (2023). Why regulators in Canada and Italy are digging into ChatGPT's use of personal information. *CBC News*.

Yin, R. K. (2018). *Case study research and applications : design and methods* (Sixth edition. ed.). SAGE.