

# **Feminisms in Artificial Intelligence: Automation Tools towards a Feminist Judiciary Reform in Argentina and Mexico**

**Ivana Feldfeber**

*DataGénero*

[ivana@datagenero.org](mailto:ivana@datagenero.org)

**Yasmín Belén Quiroga**

*DataGénero*

[yasmin@datagenero.org](mailto:yasmin@datagenero.org)

**Clarissa Guevara**

*TEC de Monterrey*

[clarissaguevara162@gmail.com](mailto:clarissaguevara162@gmail.com)

**Marianela Ciolfi Felice<sup>1</sup>**

*KTH Royal Institute of Technology*

[ciolfi@kth.se](mailto:ciolfi@kth.se)

## **Abstract**

The lack of transparency in the judicial treatment of gender-based violence (GBV) against women and LGBTQ+ people in Latin America results in low report levels, mistrust in the justice system, and thus, reduced access to justice. To address this pressing issue before GBV cases become feminicides, we propose to open the data from legal rulings as a step towards a feminist judiciary reform. We identify the potential of artificial intelligence (AI) models to generate and maintain anonymised datasets for understanding GBV, supporting policy making, and further fueling feminist collectives' campaigns. In this paper, we describe our plan to create AymurAI, a semi-automated prototype that will collaborate with criminal court officials in Argentina and Mexico. From an intersectional feminist, anti-solutionist stance, this project seeks to set a precedent for the feminist design, implementation, and deployment of AI technologies from the Global South.

**KEYWORDS:** open data, open justice, gender-based violence, feminist artificial intelligence, automation.

---

<sup>1</sup> Ivana Feldfeber is the Executive Director of the *Observatorio de Datos con Perspectiva de Género DataGénero*. She is a research fellow in the *Centro de Inteligencia Artificial y Políticas Digitales*, conducting research about public policy and AI. Ivana holds a postgraduate diploma in Data Science, Automatic Learning and its Applications from the Universidad de Córdoba, Argentina.

Yasmín Belén Quiroga is a lawyer specialising in gender and open data. She is a clerk at the Criminal Court 10 in CABA (City of Buenos Aires), Argentina. Yasmín is the co-founder of DataGénero, and teaches the gender training course "Ley Micaela" for the Judiciary of CABA.

Clarissa Guevara is a feminist lawyer currently working as a gender and police institutions trainer at the UNODC (United Nations Office on Drugs and Crime) in Mexico.

Marianela Ciolfi Felice is a Human-Computer Interaction (HCI) researcher, with a PhD in Computer Science from Université Paris Saclay. She works as an Assistant Professor in Interaction Design at the KTH Royal Institute of Technology, Sweden. Her research includes the use of AI in the frame of critical feminist computing. She is also part of a network of HCI researchers from Latin America.

## 1. Introduction

Gender-based violence is a problem that runs through and affects the lives of women and LGBTIQ+ people at the global level. This type of violence -defined as harmful acts towards a person or a group of people based on their gender<sup>2</sup>- can take many shapes: physical, sexual, psychological, economical, political, among others. Within and across feminist movements there is a long history of activism for the visibilisation and the battle against this type of violence, with a focus on feminicides of cisgender women, transgender people, non-binary people, and other gender identities. And, although GBV violates human rights and fundamental liberties, its impact remains difficult to measure.

In Latin America, the absence of traceable, high-quality data regarding GBV is a pressing issue. Today, it is not possible to combine and analyse existing data across countries, or even regions of the same country, given that this information, if available at all, is often fragmented, with temporal gaps and missing variables. As a consequence, it is hard to understand the contexts and the different manifestations of GBV in society, as well as how the associated judicial processes evolve, since the few legal rulings that become accessible to the general public are those that acquire media significance.

In addition, the credibility of the judiciary is very low: the last Latinobarómetro survey<sup>3</sup>, from 2020, shows that the level of trust in the judiciary for Latin America is 25%. As identified by Quiroga and Mandolesi (2020), *"there is a justified perception that the judicial system suffers from a big disconnect with the people that constitute the community"*<sup>4</sup>. Our project centres specifically in the cases of Argentina, which yielded a level of trust in the judiciary of 16%, and of Mexico, which got 24%. Open justice initiatives in these countries, thus, have the challenge of providing transparency as a step to bring the power of the State closer to the population.

The particularity of our project lies in that it puts the focus on the GBV cases that make it to the judiciary *before* becoming feminicides -injuries, threats, harrassment, sexual abuse, etc.-, because we see this data as key for the State to elaborate evidence-based public policy to tackle this type of violence (García, 2021). We know that the judiciary produces a large volume of data about GBV and other types of violence, and we believe that making it widely public could

---

<sup>2</sup> The definition and types can be found on: <https://www.unwomen.org/es/what-we-do/ending-violence-against-women/faqs/types-of-violence>, last accessed on June 1st, 2022.

<sup>3</sup> <https://www.latinobarometro.org/lat.jsp>, last accessed on May 30, 2022.

<sup>4</sup> This is our translation of the original quote in Spanish, which was: *"existe una fundada percepción de que el sistema de justicia padece una gran desconexión con las personas que forman parte de la comunidad"*, Quiroga and Mandolesi (2020), p. 1.

be fruitful for the public and the private sectors, feminist organisations, academia, and the general public, due to data's transformative potential. Furthermore, we argue that granting access to legal rulings in an open format would improve the transparency of the whole judiciary, would help understanding social conflictivity around justice systems, and would foster collaborative processes and innovation involving a vast range of actors. Moreover, it would help to generate trust in the judicial processes and would contribute to building fair and egalitarian societies.

The lack of political decision in this matter has brought the judiciary to create dedicated offices to partially tackle the challenge of opening judicial data. These offices produce some reports and visualisations, but have limited access to data. Although it is not within the duties of each criminal court to produce, collect, and publish open, structured gender-sensitive data, staff from a few criminal courts have started doing this work manually as part of their daily tasks. Criminal Court N° 10 from the City of Buenos Aires (from now on Court 10), is one of these few courts that carry on open gender-sensitive data initiatives, led by the judge Pablo Casas; they have been opening their data manually since 2016. The processing of these huge volumes of data about justice could benefit from the use of AI techniques to partially automate its structuring and publication, under expert human revision. This project, then, has the goal of exploring the role of AI algorithms in the automation of GBV data opening processes in the judiciary. Our project includes prototyping a tool that collaborates in this process, and testing it with real users, to then deploy it at a bigger scale in criminal courts in Argentina and Mexico.

It is key for us to refrain from framing AI algorithms as a solution to GBV. We see them, instead, as tools to better understand it and to support interventions against it (Casas, P. and Hilaire, P. 2022). Along these lines, we do not propose to automate judicial processes that require human intervention: We are against using AI to predict or make decisions in the judiciary. This position is reflected in our methodology, which is based on intersectional feminist principles (Crenshaw, 1991), epistemological surveillance (Bourdieu et al., 2004), and an approach that revisits the correspondence between data and theoretical models (Becker, 2018).

A set of questions motivate and guide our research: How can AI algorithms help us understand the phenomenon of GBV, its subtypes and their contexts? What happens with the less visible types of violence such as the psychological or the economic? How can we use these algorithms to show how judges rule, guaranteeing the protection of people's sensitive data? We hope that starting to answer these questions through our project contributes concretely to, in the long term, facilitate access to justice to women and LGBTIQ+ people who suffer GBV, and increase people's trust in the justice system.

In the remainder of this paper, we provide more detail about our positionality as authors, we situate the project in the frame of the UN's Sustainable Development Goals and of existing work in AI for social issues, justice data and GBV. Before proposing our tool, we briefly introduce the cases of Argentina and Mexico to better understand current data workflows. Then, we describe the requirements of our prototype, we illustrate its potential use in a realistic scenario, and we detail the computational techniques that we plan to explore in order to implement it. We document the methodology that we will employ, and we present a preliminary timeline that shows concretely how it is possible to carry out this project. Finally, we identify risks, limitations and opportunities of our approach, and we offer a conclusion about its feasibility and urgency.

### 1.1. Authors' positionality

The authors of this paper are four Latin American women that self-identify as intersectional feminists, based in the Global South (Argentina and Mexico) and in the Global North (Sweden), performing work and volunteer tasks in a variety of contexts (education, research, and NGOs -mostly DataGénero<sup>5</sup>). We subscribe to the feminist epistemological tradition initiated by Haraway (1991) -as we share a way of generating scientific knowledge that deviates from an androcentric universal- and taken up, in Latin America, by Maffia (2007) and by Maffia and Suárez Tomé (2021), among others. We strive for feminist research inspired by Blazquez Graf, Flores Palacios and Ríos Everardo (2010); and we adopt an intersectional approach (Crenshaw, 1991), with a focus on the triad race-social class-gender as a composite lens to analyse social inequities.

We name ourselves and our project as *feminisms* in artificial intelligence, in plural, because we are aware that the term *feminism* represents a range of very diverse fights for the people that take part in the movements. We wish to clarify that no life experience is more important than any other, nor should get more recognition; in this sense, we include queer feminism, black and indigenous feminism, feminisms against ableism and transfeminism within the feminisms that we refer to in this paper.

On top of this, we find inspiration in ideas of open government, and in particular of open justice (Castagnola, 2021). Our approach to data science is influenced by Data Feminism (D'Ignazio and Klein, 2020), which proposes to think about data, its uses and limits, guided by direct experience, and by a commitment to action and to intersectional feminism. More

---

<sup>5</sup>DataGénero civil association is a non-profit organisation in Argentina that works at the intersection of data, gender studies, and policy. Their goal is to monitor the production of data and to create sustainable data futures for our region. See more in [www.datagenero.org](http://www.datagenero.org).

specifically, we identify strongly with the principles of 'challenge power', 'consider context' and 'make labor visible'. Our position regarding the use of AI in societal issues is anti-solutionist, meaning that with this project we do not expect that AI-based tools "solve" the problem of GBV, given that this, besides being ethically irresponsible, would not be feasible, and because we believe that societal problems require societal solutions. We argue that suggesting that AI can *be* feminist is a misconception that may mislead those who are not knowledgeable about this topic: AI, algorithms and data itself cannot *be* feminist, but we can in fact think about feminist people using AI. Therefore, we believe that AI can play a positive role as a collaborator of human actors with deep knowledge of these societal issues, helping in the systematisation of high quality data on which to base public policy. This project aims to be a concrete step towards the decrease of GBV: Through data, legal rulings would become visibilised without losing context, without trying to predict nor control human behaviour, and without replacing, to any extent, the expertise of the criminal court workers.

Our collaborations and dialogues with a series of organisations in different environments influence and nourish our work. Regarding data and justice, and through DataGénero, we link up with: the Criminal Court 10 in CABA; Equis Justicia Para Las Mujeres; Intersecta; and Mumalá<sup>6</sup>. These relationships allow us to address the complexity of open, gender-sensitive justice from an intersectional, Latin American point of view. Further, the Criminal Court 10 will serve as a pilot case to design with, deploy, and test the tool that we propose. In the frame of this paper, we have the support of A+ Alliance, Women at the table, the Instituto Tecnológico de Monterrey and the Instituto Tecnológico de Costa Rica<sup>7</sup>. These organisations, through the <fAIR> (Feminist AI Research) network selected our proposal to advance feminist causes. We are also working with several partners with knowledge in Software Development, Natural Language Processing and Machine Learning, such as Cambá co-op, Collective AI society and Humai Institute<sup>8</sup>.

## 1.2. Affiliation with the UN Sustainable Development Goals

Our prototype fits the United Nations 2030 Agenda for Sustainable Development, which promotes just, peaceful and inclusive societies (SDG 16, Peace and Justice). In the context of

---

<sup>6</sup> Criminal Court 10: <https://twitter.com/jpcyf10>; Equis Justicia para las Mujeres: <https://equis.org.mx/>; Intersecta: <https://www.intersecta.org/>; Mumalá: <https://www.mumala.ar/>, last accessed on May 31, 2022.

<sup>7</sup> A+ Alliance: <https://aplusalliance.org/>; Women at the table: <https://www.linkedin.com/company/women-at-the-table/?originalSubdomain=ch>; Instituto Tecnológico de Monterrey: <https://tec.mx/es>; Instituto Tecnológico de Costa Rica: <https://www.tec.ac.cr/>, last accessed on May 31, 2022.

<sup>8</sup> Cambá co-op <https://camba.coop/>; Collective AI society <https://collectiveai.io/>; Humai Institute <https://ihum.ai/>

our work, the principle of open justice contributes to building effective and transparent accountable institutions at all levels and to ensuring public access to information and the protection of fundamental freedoms, in accordance with national laws and international agreements. Targets 3 and 7 under Goal 16 are particularly important in terms of access to justice and the rule of law. Target 3 calls on States to promote the rule of law at the national and international levels and to ensure equal access to justice for all. Goal 7 calls for ensuring inclusive, participatory and representative decision-making at all levels that is responsive to societal needs. The Judiciary is essential to the achievement of these goals and targets. Indeed, it is through this branch and its dependencies that measures can be implemented that contribute not only to Goal 16 but also to Goals 5 (Gender equality and empowerment of women) and 10 (Reduce inequalities between and within countries).

## **2. Background and Related Work**

### **2.1 AI and Societal Issues**

Scholars in Human-Computer Interaction, Science and Technology Studies and other fields have warned about applying technology to societal issues with what we can call a *techno-solutionist* approach, i.e. an approach that poses technology as a self-contained 'solution' to complex human problems (see, e.g. Blythe, 2017; Broussard, 2019; Greene, 2021; Irani, 2018). When it comes to AI, examples of biases, lack of accountability and fairness, errors, risks and harm abound in the literature, including errors in AI-made predictions having serious consequences for public health (Lazer et al., 2014); biases that produce racist results, stemming partly from lack of diversity in developer teams (Buolamwini and Gebru, 2018; see also the initiative Algorithmic justice league<sup>9</sup>); surveillance of marginalised populations (Benjamin, 2019), and so on. We are in line with this critique, as we firmly believe that technologies that aim at addressing societal problems should either do so by acknowledging and attacking the causes of inequality first, or refrain from intervening.

Our project seeks to effect change in the problem of GBV from a feminist, anti-technosolutionist perspective, which we expect to be transformative. In this sense, scholars and activists have started showing how technology, and in particular AI and data science can be indeed used as tools for common good and in particular for strengthening grassroots movements and bottom up approaches (a set of curated examples can be found in D'Ignazio and Klein, 2020); have discussed synergies between activism and academia in the context of social justice (de Castro Leal et al., 2021); and have outlined positive roles that computing could play in social

---

<sup>9</sup> <https://www.ajl.org>, last accessed on May 30, 2022

change (Abebe et al., 2020), highlighting the crucial place of feminist thinking in participation (Bardzell, 2018; Lindtner et al., 2016). The work of Savage and colleagues at the Civic AI Lab<sup>10</sup> about the future of work using AI as a collaborator of workers is an example of such alternative use cases, together with Vincenzi et al.'s (2021) approach to designing assistive AI technologies for navigation that augment the interaction of people with different abilities through interdependence, rather than posing independent navigation as a problem to fix, and AI as the solution. Aligning with Data Feminism principles, we foreground in our project the importance of questioning power dynamics, as well as counting with high quality data, an aspect still undervalued in data science and AI/ML (Sambasivan et al., 2021), in order to start using data as a political tool to base public policy on, rather than as a straightforward solution.

### **2.1.1 AI and Justice Data**

Legal rulings and other judicial documents constitute an attractive context to researchers and practitioners in the area of AI due to the big volume of data that is produced and its characteristics, including its implications for democracy. We differentiate our project from any initiative attempting to predict crime, to identify individuals in 'risk' of committing a crime, to automatise judicial decision making, and to surveil or control people's behaviours<sup>11</sup>. Other works propose to automatise legal indexing, such as that of Cumyn et al. (2019), who applied facets to a database of legal rulings from Quebec. In Latin America, and more specifically in Brazil, a number of researchers have explored the use of Named Entity Recognition (NER) in legal documents written in Brazilian Portuguese. NER refers to the task of identifying and classifying named entities in unstructured text. For example, Albuquerque et al. (2022) produced a corpus of legal documents in the frame of a project to increase transparency in the judiciary; and Zanzuz and Rigo (2022) trained BERT models with judicial data to improve legal NER, created a prototype for users in the judiciary to test the effectiveness of the model that achieved good quality results, and made the models and the prototype available publicly. BERT is a model that stands for Bidirectional Encoder Representations from Transformers. This is a transformer-based machine learning (ML) technique for natural language processing (NLP) pre-training developed by Google<sup>12</sup> in 2018.

Although these projects indeed go towards a more situated use of NLP that does not depend on models trained with English-based data, our project implies structuring the legal

---

<sup>10</sup> <https://civicai.khoury.northeastern.edu/>, last accessed on May 30, 2022.

<sup>11</sup> Following citational justice practices, we avoid giving visibility to such projects and choose not to cite them in this paper.

<sup>12</sup> [https://en.wikipedia.org/wiki/BERT\\_\(language\\_model\)](https://en.wikipedia.org/wiki/BERT_(language_model)) last access on June 2nd, 2022.



rulings' data and making it available in a way where the AI collaborates with human experts along the process, and not just when evaluating the prototype. For this reason, we see the work of Leavy et al. (2021) as inspiring for our case. These authors proposed an approach to data curation in AI that is grounded in feminist epistemology, critical race theories, and Data Feminism, with the goal of assessing matters of power in AI sociotechnical systems -similar in this dimension to Miceli et al.'s call (2022) to examine power rather than bias in the context of data production in ML.

### **2.1.2 Text document processing in Spanish**

In 2019, Jorge Pérez from the Milenio Fundamentos de los Datos Institute (IMFD) and at the University of Chile, developed BETO, a BERT model trained on a big Spanish corpus<sup>13</sup>. BETO is of size similar to a BERT-Base and can perform the same General Language Understanding Evaluation (GLUE) task set as BERT (Perez, et al, 2020), including NER tasks. There are precedents of NER trained with BERT in Spanish (Vunikili et al., 2020) and with word representations and random conditional fields (Zea et al., 2016) that yielded promising results. There is also research conducted by MuckRock, La Nación Data, Ojo Público and the Centro Latinoamericano de Investigación Periodística (CLIP), whose purpose is to extract unstructured information linked to public contracting in Argentina<sup>14</sup>. This research can serve as a precedent for creating prototypes that work in Spanish, since it deals with long documents whose information needs to be extracted and structured. The main caveat is that the AI models it implements belong to Google and AWS, so we do not wish to replicate them.

In summary, there is some promising work in this area and a variety of teams are putting considerable effort into generating better models in our language to analyse and better understand our own data. But, we know that counting with high quality data in the first place is crucial, and that is precisely what is missing at the moment.

## **2.2 Initiatives on visibilisation of GBV in Latin America**

Faced with the lack of official statistics in Latin America, individual women and women organisations made the decision, in recent years, to keep a record of feminicides published in digital and printed media, with the goals of giving visibility to the problem of GBV in their country and of sensitising society and public officials about this topic. On this path, the "La Casa del

<sup>13</sup> <https://github.com/dccuchile/beto> last access on June 2nd, 2022.

<sup>14</sup> <https://www.muckrock.com/news/archives/2021/nov/30/testing-two-named-entity-recognition-models-spanish/>, last access on May 30, 2022.



Encuentro" civil association<sup>15</sup> produced in 2008 their first report on feminicides in Argentina (entitled "Informe de Femicidios en Argentina"), analysing cases of women that, according to information acquired solely through media, were murdered in the frame of sexist violence. These periodically disseminated reports not only give a name and a face to the victims, but they also contribute significantly to start grasping the dimension of the most dramatic aspect of *machista* violence in its most extreme manifestation. Similarly, in Mexico, María Salguero created an interactive map of feminicides<sup>16</sup> committed since 2016, by gathering data from media sources, such as the victim's age, type of murder, crime scene, and so on.

More recently, activists and scholars with interest in Latin America further consolidated initiatives around visibilising and effecting change in the context of femicide using technology, and in particular machine learning. For example, "Datos contra el feminicidio" ("Data against femicide"), started by D'Ignazio, Fumega and Suárez Val in 2019<sup>17</sup>, provided automatised tools for civil organisations to collect data on feminicides from the media. One of such tools, the Highlighter, is a Google Chrome plug-in that highlights news articles about feminicides to aid in case recording; another tool, the Email Alert system, allows the user to configure filters of interest in order to receive relevant messages about femicide cases.

Although these initiatives are all fundamental contributions to show and demonstrate the severity of the problem and its societal implications, we are convinced that we need to understand how the GBV manifests in all its types (not just feminicides) and modalities (i.e. in domestic, work, institutional spheres, etc.), as a key milestone in the design of public policy towards the prevention, eradication and sanction of this type of violence.

### **2.3 Background: Judicial power and GBV in Argentina and Mexico**

Before describing our proposed prototype, we need to provide more context about the current situation of justice data on GBV in Argentina and Mexico. Today, the judiciary in Latin America use case management systems, and store a large volume of data. In federal countries, such as Argentina and Mexico, each municipality, state or province has its own system, with different characteristics, developed and maintained by different people (sometimes the State itself, sometimes private entities). Not only do they store the data following idiosyncratic rules, but none of the judiciary branches in these countries publish the data of the cases that they process. Even though there are offices dedicated to justice statistics, they do not publish gender-sensitive data, nor data fulfilling the Open Data Charter (2015). In addition to the

---

<sup>15</sup> <http://www.lacasadelencuentro.org/femicidios02.html>, last access on May 30, 2022.

<sup>16</sup> <https://femicidiosmx.crowdmap.com/>, last accessed May 30, 2022.

<sup>17</sup> <https://datoscontrafeminicidio.net/>, last accessed May 30, 2022.

heterogeneity of systems and the lack of statistics, justice-related organisations, given their nature, intervene post-facto. It is precisely because of these reasons that we need to know how the people in charge of imparting justice solve the cases that happen in a context of GBV, as a first step towards prevention. Below we briefly describe the current workflow of justice data in two cases from Argentina and Mexico.

### **2.3.1. The Argentine Case: Courts of the City of Buenos Aires**

The City of Buenos Aires courts are currently using a case management system named Expediente Judicial Electronico (EJE). This system was developed by UNITECH<sup>18</sup>. The case file timeline can be described as follows:

1. The Court receives a case through the EJE system.
2. The case is processed through a series of procedural steps, which are registered in the system by date of electronic signature (either by the public official or the judge).
3. The judge resolves a procedural situation and/or the petition brought to them.
4. The decision is uploaded into the system and it is labelled as a "legal ruling". This way, the legal ruling is incorporated into a judgement electronic book within the EJE system.

### **2.3.2 A database built and maintained by the Criminal Court 10 of the City of Buenos Aires**

The Criminal Court 10 of the City of Buenos Aires promotes, designs and enables the application of open justice policies through the use of its own public database<sup>19</sup>. This database is maintained by the people who work in the court, including Yasmín, co-author of this paper. The database has around five thousand anonymised legal rulings dated from August 2016 onwards, including a considerable number of GBV cases. It contains 64 categories with detailed information on each legal ruling, like the type of violence suffered by the victim in each case, in line with the No. 26.485 Argentina Law<sup>20</sup>. The database also includes contextual data (e.g. socio-economic variables of the people involved in the conflict, whether the defendant has

<sup>18</sup> See in <https://www.unitech-corp.com/poder-judicial/>, last accessed June 13, 2022.

<sup>19</sup> The database can be explored in the following link [https://docs.google.com/spreadsheets/d/1uAi-Yfq-rJl\\_cqQaVe9Fv1rLIBJcmEtDpUB0NTOrLAs/edit#gid=625331269](https://docs.google.com/spreadsheets/d/1uAi-Yfq-rJl_cqQaVe9Fv1rLIBJcmEtDpUB0NTOrLAs/edit#gid=625331269), last accessed May 31, 2022.

<sup>20</sup> <https://www.argentina.gob.ar/normativa/nacional/ley-26485-152155/actualizacion>, last accessed June 12, 2022.

children with the victim, the phrases used during the aggressions, etc.). Yasmín and the employees of Court 10 use different tools to maintain the database. For example they use a tool to anonymize legal rulings called IA2<sup>21</sup>. These tools are used in the daily work of the employees as follows:

*Once the legal ruling is signed by the Judge Pablo Casas, Yasmín retrieves and downloads the .docx file from the EJE system. Then, she uploads the legal ruling.docx file into the IA2 tool, which anonymises it, i.e. replaces the personal data of the persons involved in the process and assigns different labels (e.g. the name "Laura" with the word "<NAME>"). Once the legal ruling document is anonymised, Yasmín adds each field related to the case into the Court's database, using one row for each ruling and doing the process manually. Much of the data that is loaded into the database is not found within the legal ruling itself, so Yasmín must search for such data in the electronic file, yet another manual process that adds approximately 5 minutes to the processing of each ruling. The last stage involves uploading the anonymised legal ruling to a public Google Drive folder, and then pasting the link on a field in the corresponding row of the database. Yasmín has to manually repeat these steps for each legal ruling to be published.*

### **2.3.3 The Mexican case**

In July 2020, Mexico reformed its “Ley General de Transparencia y Acceso a la Información Pública” (the Transparency and Freedom of Information law). This law established the grounds, principles and procedures to guarantee the right of access to information held by any public authority, agency or organisation, including justice institutions.

In line with the mentioned reform, some justice institutions began to design and implement platforms in which they could make legal rulings available to the public. Some of these platforms included technological tools to anonymise personal data. By 2019, except for the judiciary of the State of Nuevo León and the State of Mexico, no other states had implemented any technological tool to comply with the provisions of the law (Equis Justicia para las Mujeres, 2019), and to our knowledge, the situation has not improved in the last two years.

The Judicial Branch of the State of Nuevo León, through its IT Department, designed and implemented a technological tool called “Mis Aplicaciones” (“My applications”). This tool enables the production of public versions of legal rulings, produced in a semi-automated way:

---

<sup>21</sup> IA2 is a desktop application developed in collaboration with ILDA, <https://idatosabiertos.org/about-ilda/>, in 2021. Open code: <https://github.com/instituciones-abiertas/ia2-desktop-app>, last accessed May 31, 2022.

- 1) Once the judge signs a legal ruling, the clerks are in charge of managing the tool, selecting the personal data from the legal ruling and replacing it with 10 asterisks marks.
- 2) Prior to publication, the Transparency Department may suggest deleting or anonymising more data.
- 3) The clerks look into the Transparency Department suggestions (if any), and they accept or reject them. Then, the legal ruling free from personal data, is published on the website of the Nuevo León judiciary, in the “Sentencias Públicas”<sup>22</sup> (Public Judgements) section.

Regardless of the current situation of legal rulings availability and publication by Mexican states, the mentioned law states that it is the obligation of the judiciary to publish *those legal rulings that are of public interest to society* -those whose disclosure is useful for the public to understand the activities carried out by the judiciary. In this scenario, only those legal rulings that the judge considers to be of public interest are published. We wish to highlight this reality as problematic, as the judges can potentially leave aside the publication of legal rulings made within GBV cases if they do not deem them to be of public interest.

### 3. Proposed prototype

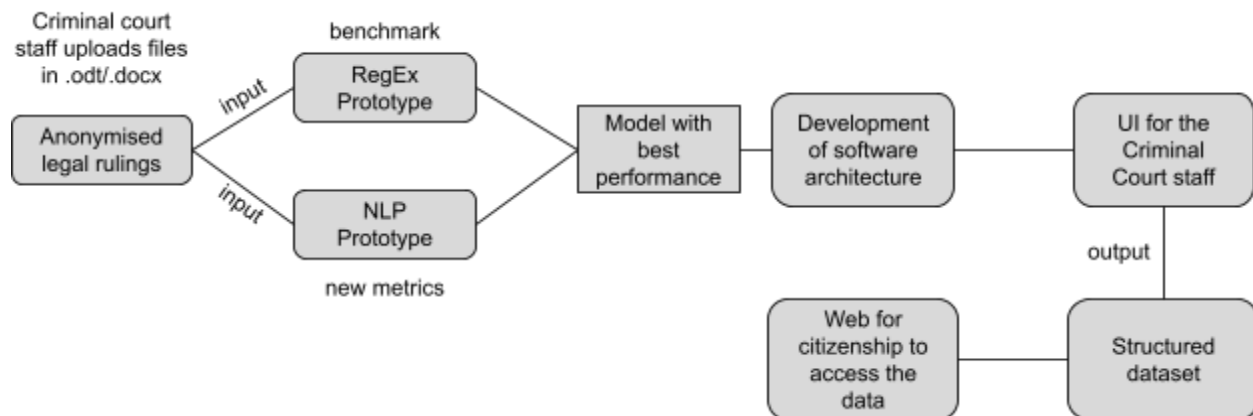


Figure 1

In this project we propose the creation of a prototype called *AymurAI* -referring to the Quechua word *aymuray* or *aymoray*, whose meaning relates to harvesting times<sup>23</sup>-, for use in

<sup>22</sup> <http://www.pjenl.gob.mx/SentenciasPublicas/>, last accessed June 5, 2022.

<sup>23</sup> <https://es.wikipedia.org/wiki/Aymoray>, last accessed June, 2022.

criminal courts in CABA and in Mexico. This prototype will employ AI techniques to partially automate the publication and maintenance of open data from the judiciary in GBV cases. Our proposal is to generate a tool that can easily identify the relevant information in a text document and extract it as a structured dataset. One of the main challenges when working with text documents is to convert unstructured information into a structured dataset, given that this task can be extremely time consuming if done manually. In spite of this functionality being seemingly available online, we are convinced that producing our own tools is the best path, since these websites are often paid services, can stop working without notice, lack clear data protection or security policies, and we believe that in the Latin American context, it is important to avoid depending on private companies when it comes to creating technology. We should also take into account that the format of the files many times doesn't match with the current website's requirements.

Our goal is that AymurAI can be used by anyone at the criminal court whose task is to keep a record of the legal rulings signed by a judge. It will not be required that the user has additional knowledge about AI or algorithms, but it would be beneficial that the involved staff receive training about the use of the tool and about the importance of opening gender-sensitive justice data. In the following section we present the requirements that the tool must fulfil, a scenario illustrating its use, and more detail about the potential AI models to explore.

### **3.1. Requirements**

By interviewing Yasmín, who has extensive expertise in open justice data as part of her daily work in Criminal Court 10 (Argentina), and in conversation with Luis Fernando (Equis Justicia para las Mujeres) from Mexico, we elicited and refined a set of requirements and design constraints that the prototype should fulfil. As we adopt a user-centred, exploratory methodology (see section 4 for more detail), we recognise that these requirements are fluid and will evolve with the iterations of the design process. We intend to use them as a guide to structure our work and to estimate the resources needed for its completion, while being open to question them along the way.

Functional requirements include: offering a user interface where a judiciary employee can login and load legal rulings; anonymising sensible data in each legal ruling; giving the user the possibility of approving/correcting the anonymed output; updating a public database with the extracted anonymised data; providing the user with a newly generated file containing the anonymised legal ruling document, also publicly available online.

Nonfunctional requirements include: security (only authorised judiciary employees should be able to log in to the system; a desktop application is preferred over a website to minimise risks of attacks); availability (the prototype should not be overly dependent on internet connection nor on its speed, given the unequal levels of infrastructure of the judiciary branches); usability and user experience (the user interface should provide a good user experience, prioritising learnability, as well as following classic guidelines such as recognition over recall). Performance, although desirable, is not crucial, as the processing tasks can be done in the background and the database can be updated along the day (for example, when internet connection is available). Scalability is not needed at the level of the desktop application, but at the level of the public database, which would be potentially updated by several judiciary branches at the same time.

These requirements exist within the frame of certain design constraints that we have so far identified: To be feasible, the new workflow that *AymurAI* would bring about cannot radically change, at least in the short term, the way in which data is produced about legal rulings. For example, the input to *AymurAI* will have to be a set of Microsoft Word files (ODT or DOCX format), each with one legal ruling.

### 3.2. Design Scenario

Through a vignette of a design scenario based on the requirements detailed above, we introduce *AymurAI*. We imagine, then, that *AymurAI* is being widely used in CABA's criminal courts:

*All the criminal courts in CABA apply open justice policies through a public database, updated automatically by AymurAI, a safe tool whose results are checked by the staff of each criminal court prior to their publication.*

*The database stores anonymised legal rulings, including cases classified by judges as GBV, as well as other types of cases. The legal rulings contain detailed information about the types and subtypes of violence that were present in each case and their modality, as well as socioeconomic data of the individuals involved, and other contextual data (for example, children in common, frequency of violence episodes, crime scene(s), among others). The following scenario exemplifies how Yasmín, from Criminal Court 10, uses AymurAI in her daily tasks:*

*In the morning, Yasmín logs in to AymurAI, a desktop application installed in the computers of the criminal court, which are connected to a server in the Consejo de la Magistratura. She sees that all the legal rulings signed the previous day by the judge are*

*available in ODT format, ready to be processed and then published. Yasmín loads the legal rulings into AymurAI by simply choosing the folder in which they are stored. AymurAI processes them one by one.*

*For each legal ruling, AymurAI recognises a set of entities and proposes to anonymise some of them, replacing them with meaningful labels. For example, the address of the scene under investigation gets replaced with "<ADDRESS>". Anonymised entities are shown on screen so that Yasmín checks them, accepting them (correct cases) or rejecting them (false positives). She can also signal to AymurAI if some entity was not captured (false negatives), given that this could have serious consequences in case of becoming public. Once she is done with the first legal ruling, AymurAI shows Yasmín the second one, and so on until they are all processed.*

*Yasmín's tasks are completed. AymurAI lets her know that the structured data and the anonymised legal rulings will be reflected in the database in some minutes. The gender-sensitive database, which gets updated thanks to AymurAI, is open and thus is available so that any interested party -citizen, state body, civil society organisation- can consult it, through a website from the Consejo de la Magistratura.*

### **3.3. Models**

Our prototype will be developed in two stages. The first stage aims to generate a benchmark with simple rule-based AI techniques using regular expressions. Our target dataset has around 50 columns, these are the current columns that Court 10 decided to open to the general public, minus the ones regarding the appeal court, because it is out of our scope. The goal of this first model will be, through such control structures, to identify the data needed to conform the dataset, and successfully extract them from the text. The second model will implement more complex machine learning models using NLP to identify the entities corresponding to the target dataset fields.

#### **3.3.1 Regular Expressions**

There is a precedent of the *regex*<sup>24</sup> technique applied to justice data in Argentina: through regular expressions and programming rules, the IA2 prototype (section 2.3.2) was built by the Cambá co-op in partnership with ILDA and Criminal Court 10, which currently detects sensitive information in legal rulings, including fields such as name, address, date, number of case, and replaces them with a generic label. This technique could be further enhanced to

---

<sup>24</sup> <https://regexr.com/> last accessed June 5, 2022.



detect the data that we need in the output. To extract the necessary information from these files and dump it into a structured dataset, we propose to use a set of Python libraries. On the one hand, there are relevant libraries for text processing such as *python-docx*<sup>25</sup> which quickly extracts the content of Word documents into paragraphs. Then with *textract*<sup>26</sup> and regex we can generate simple rules to find, mark and extract certain categories from the texts.

Once this model is working, we will generate the metrics that will serve as a benchmark for comparisons within the second stage.

### 3.3.2 Natural Language Processing

The second modelling stage will involve the implementation of more complex Machine Learning models, using NER, the most basic NLP technique that seeks to extract entities in a text. This technique can automatically scan entire articles, identify key entities and classify them into predefined categories. In order to work with NER in Spanish it might be necessary to manually tag documents with the target categories, in order to have a large enough corpus to yield good results. For this purpose, we will also use NLP libraries, such as SpaCy<sup>27</sup> and HuggingFace<sup>28</sup> (within the latter there are many models trained with Spanish corpora that can be tried with our own).

Once we have the models running, we can compare them with our benchmark and through an analysis of factors involving the comparison of: classification accuracy, explainability, environmental impact, inference time, maintenance, and replicability. We will choose the one that works best taking trade-offs into account, and then build the software architecture according to the design scenario developed in section 3.2.

---

<sup>25</sup> <https://python-docx.readthedocs.io/en/latest> last accessed June 5, 2022.

<sup>26</sup> [https://textract.readthedocs.io/en/stable/python\\_package.html](https://textract.readthedocs.io/en/stable/python_package.html) last accessed June 6, 2022.

<sup>27</sup> <https://spacy.io/> last accessed Jun 6th, 2022

<sup>28</sup> <https://huggingface.co/> last accessed Jun 6th, 2022

#### 4. Method and Plan

Tasks	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
1. Design of requirements and specs with the development team	X					
2. Labelling documents	X					
3. Developing regex model and metrics	X	X				
4. Developing machine learning model and metrics	X	X	X			
5. Evaluating and comparing models' performance			X	X		
6. Accounting for biases and inaccuracies		X	X	X	X	X
7. Developing prototype			X	X	X	
8. Testing for usability					X	X

**Figure 2**

We plan to combine the theoretical influences of this project (detailed in sections 1.1 and 2) with its practical requirements (section 3.1), through a feminist data science approach, strongly guided by Data Feminism (D'Ignazio and Klein, 2020), Human-Centered Data Science (Aragon et al., 2022) and state-of-the-art guidelines on Human-Centred Machine Learning (Chancellor, 2018). We will address our research questions with an exploratory, iterative spirit centred on users (leveraging on and prioritising the knowledge of the criminal court staff and the local feminist organisations that work with justice data), with a solid presence of ethical questionings along the whole design and development process. For example, before deciding to add a new feature, the team will discuss possible misuses of such a feature, pondering benefits over risks, assessing whether risks could be mitigated, or whether the feature should be discarded. Design decisions will be documented and published via Github, rendering AymurAI's features traceable over time, and making the process auditable. The team that will prototype the tool is composed of this paper's authors plus more members of DataGénero and experts in software development and machine learning/NLP, who are members of technology co-ops following the principles of open knowledge and free software<sup>29</sup>.

<sup>29</sup> Potential actors in this context include the Cambá co-op <https://camba.coop/> and Collective AI society <https://collectiveai.io/>.

Figure 2 shows a preliminary project timeline, starting with the first exploratory phase with the development team to design the requirements and specifications of the project to start working and at the same time there will be human-made labelling of legal rulings. During the first month we will start the development of the regex model and its metrics and of the machine learning model and its metrics. Given that we have 50 different categories, maybe we will need metrics for many of them, since they have different kinds of data inside of them. After comparing and evaluating the models, looking for the one with the best performance, we will start building the prototype. The evaluation of the models, after defining our metrics will be addressed during the third month of our project in collaboration with the development team. During the whole process, we will be looking for biases and inaccuracies. Finally, we will test the tool's usability to implement further refinements. This prototype can be tested in Criminal Court 10 that will work as a pilot in our project. As shown in the figure, to be feasible in six months, the project will prioritise the automatic creation of a human-approved, open database with anonymised information about legal rulings (as a first step, coming only from Criminal Court 10 in CABA), that can be later accessed by anyone. The long-term maintenance of this database, as well as the creation of interactive visualisations, infographics, or statistics on the most frequent queries, are out of the scope of this prototype and would be a certainly important part of future work in the context of linkage with the broader community.

## 5. Risks, Limitations and Opportunities

In conversations among co-authors, we have identified a set of preliminary risks and strategies to mitigate them or to signal their existence when this is not possible. Risk assessment, however, will be done in a continuous way along the project, as described in section 4. An evident source of risk is that of bias in the data about GBV. For example, the judiciary branch in CABA only has criminal jurisdiction in certain types of crimes<sup>30</sup>. It is important, then, to bear in mind that the absence of other types of crimes in the resulting datasets will be due to this fact. Providing context about the nature and limits of the available data in the dataset would be needed to ensure that citizens (or the state itself) consulting it do

---

<sup>30</sup> It has competence to hear cases involving the offences set forth in the Criminal Code which have been transferred to the jurisdiction of the courts of the City of Buenos Aires (under Laws N° 25752, 26357, and 26702: the offences transferred are: injuries, duel, abuse of firearms, breaking and entering, arson and other forms of property destruction, possession, carrying and provisioning of war weapons, hindering or preventing contact, criminalisation of discriminatory practices, offenses and contraventions related to sports and sports events, crimes against the public administration, counterfeiting and forgery, and crimes related to the jurisdiction of the local public administration), contraventions -such as those related to conflicts between neighbours- and misdemeanors -which are usually violations to commercial and traffic regulations.

not incorrectly assume that other types of crimes simply do not occur in the city. In addition, each court in CABA gets assigned the cases that happen within its geographical area: As each area has its own socioeconomic variables -exhibiting a high degree of heterogeneity in some cases-, ulterior data analysis should take this into account. We will use the awareness of these risks of bias from the beginning of the process, in order to think and dialogue with the data on the way. Another source of risk is data de-anonymisation. For example, in towns with few inhabitants, additional data obfuscation could be needed to make sure that people cannot be identified, given that the number of specific types of crime in a given temporal window is likely to be small or even unique. We will address this during the prototyping phase as well as during the deployment phase (which is out of the scope of this paper), in which we plan to deploy the prototype as a pilot in CABA's Criminal Court 10. Later on, provided that further deployment resources are available, we will seek to implement a version of the prototype, or at least formalised lessons learnt in a Mexican court, as part of the "Pacto por una Justicia Abierta con Perspectiva de Género"<sup>31</sup>.

As mentioned earlier, federalism poses challenges to the adoption of a generalised approach, given the idiosyncrasies of each court and the ecosystem of existing hybrid (digital and analogue) systems, together with -in the view of the authors- a reticence in the judiciary to embrace change. We choose, in this project, to take a bottom up approach where we focus first on two particular cases, heavily informed by the workers in these courts and our own cultural and political knowledge of the territories. Following our feminist stance, we aim at achieving *transferability* of the new knowledge that will be produced through the prototype in use, rather than pretending to generate a universal solution that fits all the possible situations without the need of local adaptations. In fact, we expect that at the technical level, the prototype could be designed to be malleable enough to address a wide range of courts. Instead, it is the political dimension of it that would be the main limitation to its adoption. Still, this project is set to be a strong, concrete exemplar of open justice advocacy that will hopefully effect change.

This project has the potential of going well beyond the main goals that we have detailed in this paper. It opens up a set of research and outreach opportunities, including long deployment periods in the pilot courts, where in-the-wild studies would shed further light on how justice is imparted and how the prototype affects the work practices at the courts and generates new ones; adding new courts progressively, which would imply making the prototype more versatile; as well as running user studies about the prototype's user experience, not just to improve it but to inform the design of justice-related sociotechnical systems.

---

<sup>31</sup> <https://pactojusticiaabierta.mx/> last accessed June 13, 2022.

Another logical follow up of this project, where the link to the community can be strengthened, is that of exploring how the resulting open database could be consulted by citizens and by the state: How to provide context about the data in the dataset to, for example, acknowledge and foreground existing biases in the data? How to leverage the data's communication power through interactive visualisations? How can information visualisation techniques be used to direct people's attention to important yet less media-attractive aspects of justice? Are the results used to elaborate public policy, or to support activism, and how?

At a longer term, and predicting that the volume of digitally available data about justice will grow at an ever higher rate, we envision that an office in each court could be dedicated to work with the legal rulings in close collaboration with AymurAI, taking care of, for example, keeping track of each ruling's history and publishing it, to ensure traceability. A wider adoption of our prototype, then, would not make jobs obsolete but rather create opportunities to leverage digitalisation and take advantage of AI in a feminist way, where human expertise is highly valued, and where it is the AI the one who is in the loop, as a partner or a tool rather than as a gatekeeper or a decision maker (Ciolfi Felice et al. 2021).

## 6. Conclusions

In this paper we advocate for the structuring and publication of judicial data on gender-based violence in Latin America -starting with Argentina and Mexico-, with the long-term goals of facilitating the access to justice to women and LGBTIQ+ victims of GBV, and of increasing the general public's trust in justice. We have argued that it is crucial to focus on this type of violence so that governments can widely adopt evidence-based public policies that prevent and eradicate it *before* these cases may become feminicides. We seek to enhance the debate around the need of a feminist judiciary reform, unveiling the obscure decision-making of judges (overwhelmingly male, cisgender, white, heterosexual and wealthy), to produce rulings and decisions that can protect and create better lives for women and LGBTIQ+ individuals around the world.

We have identified the potential of AI algorithms -in particular, regular expressions and some NLP tasks such as NER with SpaCy and HuggingFace- to collaborate with human experts in the process of opening judicial data. Moreover, we have proposed a prototype -AymurAI- that would use AI to learn the structure of the data in legal rulings, and would create and maintain a public database with the anonymised cases. From an intersectional feminist, anti-solutionist stance, we have described the theoretical and methodological frameworks that guide this project, with an estimated timeline. Not only through the use of AymurAI in courts but also

during its feminist design and development process we expect to gather concrete insights on how AI algorithms can help us understand the phenomenon of GBV and how the judicial system addresses it in each country. While we are particularly interested in data about GBV, we commit to an intersectional approach to societal inequalities, which can make these tools applicable when examining systems of oppression involving ableism, racism, and classism, among others. Finally, this paper shows that AymurAI would constitute a grounded yet feasible example of how AI can play a positive role in the real world without replacing humans in decision making; and of how feminist principles can be used *by people* to guide and audit the process, and to visibilise human labour, rather than expecting AI algorithms to *be* feminist and to *do* the feminist work.

## REFERENCES

- Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020). Roles for computing in social change. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3351095.3372871>
- Aguiar, A., Silveira, R., Furtado, V., Pinheiro, V., & Neto, J. A. M. (2022, March). Using Topic Modeling in Classification of Brazilian Lawsuits. In *International Conference on Computational Processing of the Portuguese Language* (pp. 233-242). Springer, Cham.
- Albuquerque, H. O., Costa, R., Silvestre, G., Souza, E., da Silva, N. F., Vitório, D., ... & Oliveira, A. L. (2022, March). UlyssesNER-Br: A Corpus of Brazilian Legislative Documents for Named Entity Recognition. In *International Conference on Computational Processing of the Portuguese Language* (pp. 3-14). Springer, Cham.
- Aragon, C., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-Centered Data Science An Introduction*. The MIT Press.
- Balaam, M., Comber, R., Clarke, R. E., Windlin, C., Ståhl, A., Höök, K., & Fitzpatrick, G. (2019). Emotion work in experience-centered design. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.  
<https://doi.org/10.1145/3290605.3300832>
- Bardzell, S. (2018). Utopias of Participation. *ACM Transactions on Computer-Human Interaction*, 25(1), 1–24. <https://doi.org/10.1145/3127359>
- Becker, H. (2018). *Datos, pruebas e ideas*. Siglo XXI.
- Benjamin, R. (2020). *Race after technology: Abolitionist Tools for the new jim code*. Polity.



Blazquez Graf, N.; Flores Palacios, F. & Ríos Everardo, M. (coord).(2010). *Investigación feminista : epistemología, metodología y representaciones sociales*. UNAM, Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades.

[http://biblioteca.clacso.edu.ar/Mexico/ceiich-unam/20170428032751/pdf\\_1307.pdf](http://biblioteca.clacso.edu.ar/Mexico/ceiich-unam/20170428032751/pdf_1307.pdf)

Blythe, M. (2017). Research fiction. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3025453.3026023>

Bourdieu, P., Chamboredon, J. C. & Passeron, J. C. (2004). *El oficio del sociólogo*. Siglo XXI.

Broussard, M. (2019). *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

*Carta Iberoamericana de Gobierno Electrónico - clad*. (n.d.). Retrieved May 29, 2022, from <https://clad.org/wp-content/uploads/2020/07/Carta-Iberoamericana-de-Gobierno-Electronic-o.pdf>

Casas, P. & Hilaire, P. (2022) Cómo la Inteligencia Artificial puede expandir la Justicia Abierta y proteger los datos personales. In *Hacia una agenda global de justicia abierta para América Latina*. Editorial Jusbaire.

Castagnola, A. (2021). *Justicia Abierta En Tiempos De Covid 19. Un modelo para armar en base a las experiencias del Juzgado Penal, Contravencional y Faltas No. 10 de la Ciudad Autónoma de Buenos Aires*. PNUD.

Chancellor, S. (2022). Towards Practices for Human-Centered Machine Learning. *arXiv preprint arXiv:2203.00432*.

Ciolfi Felice, M., Fdili Alaoui, S. & Mackay, W E. (2021). Studying Choreographic Collaboration in the Wild. In: *Designing Interactive Systems Conference 2021* (DIS '21) New York, USA <https://doi.org/10.1145/3461778.3462063>

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, 43 (6), pp. 1.241-1.299.

Cumyn, M., Hudon, M., Mas, S., & Reiner, G. (2018). Towards a new approach to legal indexing using facets. *Lecture Notes in Computer Science*, 881–888.  
[https://doi.org/10.1007/978-3-319-92058-0\\_84](https://doi.org/10.1007/978-3-319-92058-0_84)

D'Ignazio , C., & F. Klein, L. (2020). *Data Feminism . Data Feminism. The MIT Press*. (MIT Press). Retrieved 2022, from <https://mitpress.mit.edu/books/data-feminism>.

D'Ignazio, C., Val, H. S., Fumega, S., Suresh, H., & Cruxên, I. (2020). Feminicide & machine learning: detecting gender-based violence to strengthen civil sector activism.

EQUIS JUSTICIA PARA LAS MUJERES. (2019). *(IN)justicia Abierta - Equis*. Retrieved May 29, 2022, from  
[https://equis.org.mx/wp-content/uploads/2019/05/INJusticia\\_Abierta.pdf](https://equis.org.mx/wp-content/uploads/2019/05/INJusticia_Abierta.pdf)

EQUIS JUSTICIA PARA LAS MUJERES. (2019). *Diagnóstico: La Enorme Opacidad de los poderes judiciales*. Equis. Retrieved May 29, 2022, from  
<https://respaldo.equis.org.mx/diagnostico-la-enorme-opacidad-de-los-poderes-judiciales/>

Fundación para la Justicia y el Estado Democrático de Derecho. (2021). *Imumi.org*. EL ACCESO A LA JUSTICIA EN MÉXICO DURANTE LA PANDEMIA DE COVID-19. Análisis

sobre la actuación del Poder Judicial de la Federación. Retrieved May 29, 2022, from <https://imumi.org/wp-content/uploads/2021/08/el-acceso-a-la-justicia-en-mexico-durante-la-pandemia-de-covid-19.pdf>

García, M. (2021). *Una alianza estratégica para la igualdad de género. Datos con perspectiva de género para el desarrollo de políticas públicas*. En Género, Estado y políticas públicas. XV Congreso Nacional de Ciencia Política. Universidad Nacional de Rosario.

Greene, D. (2021). *The Promise of Access Technology, Inequality, and the Political Economy of Hope. The Promise of Access*. The MIT Press. Retrieved 2022, from <https://mitpress.mit.edu/books/promise-access>.

Haraway, D. (1991). “Conocimientos situados: la cuestión científica en el feminismo y la perspectiva parcial”, en *Ciencia, cyborgs y mujeres. La reinvención de la naturaleza*. Madrid, Cátedra, pp. 313-346.

Irani, L. (2018). “Design thinking”: Defending Silicon Valley at the apex of Global Labor Hierarchies. *Catalyst: Feminism, Theory, Technoscience*, 4(1), 1–19. <https://doi.org/10.28968/cftt.v4i1.29638>

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, 343(6176), 1203–1205. <https://doi.org/10.1126/science.1248506>

Leal, D. de, Strohmayer, A., & Krüger, M. (2021). On activism and Academia. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3411764.3445263>

Leavy, S., Siapera, E., & O'Sullivan, B. (2021, July). Ethical Data Curation for AI: An Approach based on Feminist Epistemology and Critical Theories of Race. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 695-703).

Lindtner, S., Bardzell, S., & Bardzell, J. (2016). Reconstituting the utopian vision of making. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2858036.2858506>

Maffia, D. (2007). Epistemología feminista: La subversión semiótica de las mujeres en la ciencia. *Revista Venezolana de Estudios de la Mujer*, 12(28), 63-98.  
[http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S1316-37012007000100005&lng=es&tling=es](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1316-37012007000100005&lng=es&tling=es).

Maffía, D. & Suárez Tomé, D. (2021). Epistemología Feminista. En: Susana Gamba y Tania Diz, *Nuevo diccionario de estudios de género y feminismos*. EUDEBA

Miceli, M., Posada, J., & Yang, T. (2022). Studying up Machine Learning Data. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1–14.  
<https://doi.org/10.1145/3492853>

México Evalúa. (2022). *DIAGNÓSTICO DE IMPLEMENTACIÓN DE HERRAMIENTAS TECNOLÓGICAS en los poderes judiciales en México*. México Evalúa. Retrieved from <https://www.mexicoevalua.org/mexicoevalua/wp-content/uploads/2022/02/diagnostico-3feb-ok.pdf>

Open Data Charter. (2015). Retrived May 31, 2022, from <https://opendatacharter.net/principles-es/>

Perez, J., Cañete, J., Chaperon, G., Fuentes, R., Ho, J., Kang, H. (2020) “Spanish Pre-Trained BERT Model and Evaluation Data”. In PML4DC at ICLR 2020.

<https://users.dcc.uchile.cl/~jperez/papers/pml4dc2020.pdf>

Quiroga, Y. B., Ivana, F., García, M., Bercovich and Szulmajster, S. (2021). *¿Por qué necesitamos datos con perspectiva de género?* gad. Retrieved May 29, 2022, from

<https://gad.uy/por-que-necesitamos-datos-con-perspectiva-de-genero/>

Quiroga, Y., Mandolesi, M. A. (2020). *Apertura de Datos con Perspectiva de Género*.

<http://www.pensamientopenal.com.ar/system/files/2020/10/doctrina49536.pdf>

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021, May). “Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

Vincenzi, B., Taylor, A. S., & Stumpf, S. (2021). Interdependence in action. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–33.

<https://doi.org/10.1145/3449143>

Vunikili, R., Supriya, H. N., Marica, V. G., & Farri, O. (2020, September). Clinical NER using Spanish BERT Embeddings. In *IberLEF@ SEPLN* (pp. 505-511).

Zanuz, L., & Rigo, S. J. (2022, March). Fostering Judiciary Applications with New Fine-Tuned Models for Legal Named Entity Recognition in Portuguese. In *International Conference on Computational Processing of the Portuguese Language* (pp. 219-229). Springer, Cham.

Zea, J. L. C., Luna, J. E. O., Thorne, C., & Glavaš, G. (2016, August). Spanish NER with word representations and conditional random fields. In *Proceedings of the sixth named entity workshop* (pp. 34-40).