The disconnect between the goals of trustworthy AI for law enforcement and the EU research agenda

Sanz-Urquijo, B.; Fosch Villaronga, E.; Lopez-Belloso, M.

ORIGINAL RESEARCH

The disconnect between the goals of trustworthy AI for law enforcement and the EU research agenda

B. Sanz-Urquijo[1] · E. Fosch-Villaronga[2] · M. Lopez-Belloso[3]

Abstract

In this paper, we investigate whether AI deployment for law enforcement will enable or impede the exercise of citizens' fundamental rights by juxtaposing the promises and policy goals with the crude reality of practices, funded projects, and practicalities of law enforcement. To this end, we map the projects funded by H2020 in AI for law enforcement and juxtapose them to the goals and aims of the EU in terms of Trustworthy AI and fundamental rights. We then bring forward existing research stressing that AI implementation in sensitive domains such as defense and law enforcement does not come without drawbacks, especially regarding discrimination, surveillance, data protection, and human dignity. We thoroughly analyze and assess human-centric and socially-driven lens risks and threats of using AI factors from an ethical, legal, and societal perspective (ELSA), including organizational and gender worries.

1 Introduction

Artificial intelligence (AI) is a strategic priority on the European agenda. Europe wants to scale positions in the international arena of technological innovation by promoting human-centered Artificial Intelligence based on trust and European core values in different areas, including law enforcement [3, 21]. AI has many potential applications in law enforcement, including predictive policing [14], automated monitoring [1], (pre-) processing large amounts of data (e.g., image recognition from confiscated digital devices, police reports, or digitized cold cases) [20], finding case-relevant information to aid investigation and prosecution [31], providing more user-friendly services for civilians (e.g., with interactive forms or chatbots) [25], and generally enhancing productivity and paperless workflows.

These advancements work as a double-edged sword, however [23]. While AI could be potentially used to promote fundamental societal values that should govern, steer, and accompany police operations like human dignity, freedom, equality, solidarity, democracy, and the rule of law [9], at the same time, it challenges the values carefully guarded in existing operations and procedures [13]. It may also have implications and unintended consequences for civil society. For instance, evidence shows that face recognition systems continuously discriminate against dark-skinned people, especially women (Buolamwini and Gebru [5, 17] EC 2020). Other research points out that algorithms may exacerbate other stereotypes, including gender stereotyping or the feminization of gay males [11]. When these systems are used in national security matters, they risk exacerbating existing biases that do not reflect reality and affect the presumption of innocence, e.g., dark-skinned people are more prone to commit a crime (Skeem and Lowenkamp [24]). While the European commitment to trustworthy AI fosters civil society's belief in these tools, the civil society is not always involved in their decision-making practices, even if they are directly affected by such decisions.

✉ B. Sanz-Urquijo
borja.sanz@deusto.es

E. Fosch-Villaronga
e.fosch.villaronga@lawleideuniv.nl

M. Lopez-Belloso
mlbelloso@deusto.es

1   Faculty of Engineering, Deusto University, Bilbao, Spain

2   eLaw Center for Law and Digital Technologies, Leiden University, Leiden, The Netherlands

3   Social and Human Sciences, Deusto University, Donosti-San Sebastian, Spain

Despite the numerous pieces of literature highlighting the fundamental problems that automation causes in many sensitive application domains, the EU believes in the promise of technology and has dedicated several work programs and EU funding to the development of AI and law enforcement. In this paper, we investigate whether AI deployment for law enforcement will enable or impede the exercise of citizens' fundamental rights by juxtaposing the promises and policy goals with the crude reality of practices, funded projects, and practicalities of law enforcement. In particular, we focus on the research projects funded by the European Commission and, more specifically, how these projects support the European objectives of Trustworthy AI. To this end, we map the projects funded by H2020 in AI for law enforcement and juxtapose them to the goals and aims of the EU in terms of Trustworthy AI and fundamental rights. The projects funded by the EC research programs respond to the European Union's priorities. These priorities support a technology development model based on the EU's fundamental values, ethics, and trustworthiness, and allocating public money in sectors where its application and the protection of fundamental rights are in doubt. Therefore, we bring forward existing research stressing that AI implementation in sensitive domains such as law enforcement does not come without drawbacks, especially regarding discrimination, surveillance, data protection, and human dignity. We thoroughly analyze and assess human-centric and socially-driven lens risks and threats of using AI factors from an ethical, legal, and societal perspective (ELSA), including organizational and gender worries.

After identifying the challenges and gaps in the European Union's political and regulatory advances in this arena, we propose a comprehensive ethical-legal-societal multifold approach to assess these technologies' adverse implications for society, which is based on instruments developed by the European Commission itself but that leaves to the will of the actors their implementation (at least until the entry into force of the proposed AI Act) [18]. In concrete, we put forward ways in which the Assessment List for Trustworthy AI (ALTAI) could be implemented as a mandatory requirement before conducting these practices to provide feedback on the policy-to-practice-to-policy loophole. Overall, we argue that by providing truthful environments (e.g., co-creation spaces where all the stakeholders involved in the development of these technologies can express their opinions and try to develop the tool that covers as many points of view as possible) where discussions and challenges are facilitated, international cooperation between LEAs, law enforcement agencies, and other relevant stakeholders could be encouraged to foster solutions that account for all the voices affected by these tools [2].

2 The European digital strategy

2.1 A European take on AI policy

The draft Regulation for Artificial Intelligence presented by the EU in April 2021 (also called AI Act, 2021) [18] is the culmination of an overall strategy to develop a home-grown approach to the development of the technology that the EU initiated in 2018 with the creation of an Expert Group on Liability and New Technologies (*see* Fig. 1).

This European strategy, initiated in 2018, sought to foster European capacity and competitiveness in the AI sector, prepare and anticipate the changes that these technologies may trigger, and ensure an ethical and legal framework based on the EU's fundamental values and the EU Charter of Fundamental Rights [26].

This regulatory proposal has not been without critics [27]. Some authors focused on bureaucracy and its impact on technology developers [27]. Others focused on the limitations of such an instrument to protect citizens' rights [12] and inconsistencies with the chronology above. In its strategy to position itself in the AI market, Europe has chosen to base its approach to this technology on protecting human rights and defending the Union's fundamental values. This rationale derives from all the documents on AI, notably the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions—Building trust and confidence in human-centered artificial intelligence [7], the White Paper on AI [28] and the Communication on promoting a European approach to artificial intelligence (COM [8], p. 205 final).

This latest Communication culminating in the regulatory evolution proposes four key aspects to ensure a European approach to AI:

1. Cooperation framework to ensure the necessary conditions for the development and uptake of AI in the EU
2. Fostering ecosystems of excellence from lab to market through programs like Horizon Europe.
3. Ensuring that AI works for people and is a force for good in society. In other words, it is human-centered, sustainable, safe, secure, inclusive, accessible, and reliable.
4. Advancing in building strategic leadership in high-impact sectors such as climate change and environment, health, public sector, robotics, mobility, security and home affairs, and agriculture

The combination of these four aspects with the principles set out in the proposed regulation of AI (risk-based approach and protection of human rights) reveals certain inconsistencies in the European strategy, which seeks to base
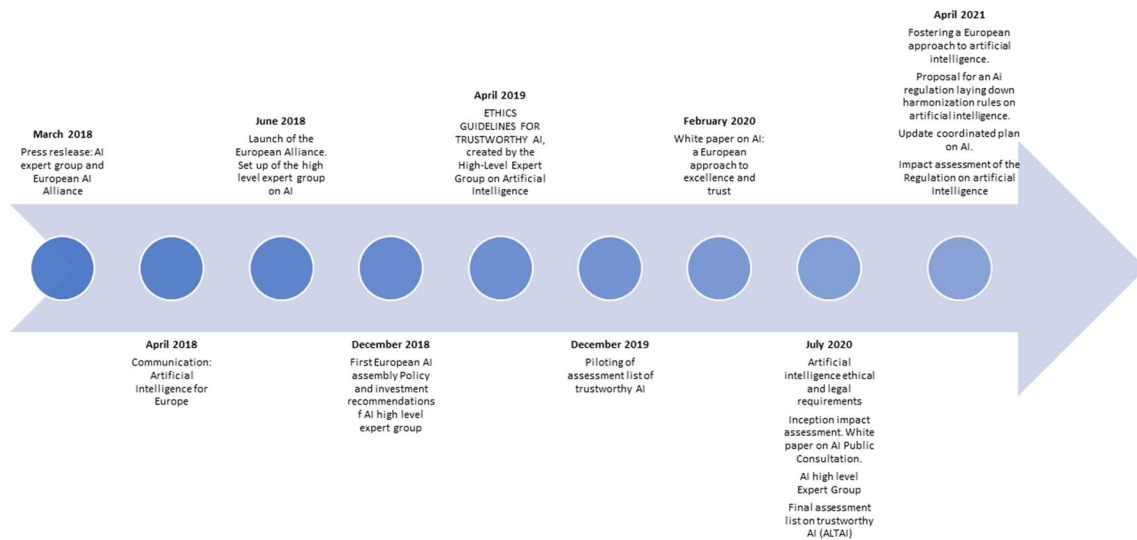
Fig. 1 European Digital Strategy chronology (See https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence)

its approach on a vision of AI structured around ethical principles and the protection of human rights. This disconnect is noticeable in its research strategy: the EU has been using the framework programs for research funding to develop technology classified as high-risk by the AI Act. Since these technologies have severe shortcomings in their vocation to protect fundamental rights, we wondered whether this was illustrative of a "Do as I say, but not as I do" approach.

2.2 A definition for AI

There are multiple definitions of AI, including many scholars suggesting that AI is neither Artificial nor Intelligent. We used the definition of the HLEG (HLEG [15, 16]) on AI which defines it as 'systems that display intelligent behavior by analyzing their environment and taking actions—with some degree of autonomy—to achieve specific goals.' This definition from the European Commission formed the basis for the subsequent proposal for the AI Act [18], which is the first-ever proposal for a legislation on AI in the EU. In 2021, The AI Act defined AI [18] as 'software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.' Until now, all the definitions referred to other disciplines (e.g., computer science, engineering). This is the first time the EU institutions defined the concept of Artificial Intelligence from a policy perspective. Since the European

Commission has the power to draft legislation, we consider this definition relevant and authoritative.

The HLEG on AI stressed that the deployment of AI should be *trustworthy*, meaning that 'as digital technology becomes an ever more central part of every aspect of people's lives, people should be able to trust it.' [10] Trust however means many things and is very contextual [4]. According to the HLEG, however, trustworthiness refers to an AI system that is ethical, lawful, and robust [15, 16]. Following the foundations for trustworthy AI as established by the HLEG, an AI system will be trustworthy an AI system supporting law enforcement should:

(a) be based on human rights and core European values;
(b) endorse the principle of non-discrimination captured in Art. 21 of the European Charter of Human Rights (EU CFR), paying particular attention to intersectional approaches such as race, gender, sex, sexual orientation, religion, political orientation, and
(c) integrate a solid ethical component and societal impact.

3 Goals and methods

Given these states of affairs, the disconnection between discourse and actions seems to be twofold: at the macro and meso levels [19]. On the one hand, the macro level (corresponding to policy goals and priorities), where priorities are set in research funding programs, do not require incorporating the values and principles ensuring the policy objectives

Table 1 Research programs and EC contributions

Program Title	TOTAL EC Budget Contribution
Access Control	4,102,467.00 €
Border crossing points topic 1: Novel mobility concepts for land border security	4,501,877.50 €
Border crossing points topic 2: Exploring new modalities in biometric-based border checks	4,981,752.50 €
Building modern rural policies on long-term visions and societal engagement	4,999,747.50 €
Ensure privacy and freedom, including in the Internet and enhance the societal, legal and ethical understanding of all areas of security, risk and management	47,246,185.38 €
EXCELLENT SCIENCE—Marie Skłodowska-Curie Actions	3,989,295.61 €
Forensics topic 1: Tools and infrastructure for the extraction, fusion, exchange and analysis of big data including cyber-offenses generated data for forensic investigation	11,992,553.25 €
Improve cyber security	3,603.831.75 €
INDUSTRIAL LEADERSHIP—Leadership in enabling and industrial technologies—Information and Communication Technologies (ICT)	1,999.951.25 €
Law Enforcement capabilities 2: Detection and analysis of terrorist-related content on the Internet	4,998,527.88 €
Maritime Border Security topic 1: radar systems for the surveillance of coastal and pre-frontier areas and in support of search and rescue operations	7,992,312.50 €
Preparatory Actions for new FET Flagships	999,250.00 €
PRIORITY 'Societal challenges	50,000.00 €
Total	101,457,752.12 €

for getting funding. On the other hand, at the meso level (the ecosystem generated by research projects), where the discourse is developed to align the proposals submitted to obtain funding, the project's specific objectives and actions do not correspond to those of the research projects. Our contribution explores the second disconnect by the following research question: Do EU-funded projects support the EU's goals for Trustworthy AI? To do so, we have combined a text-based analysis of H2020-funded projects in the area of technology and law enforcement that started in 2015 and will finish in the upcoming 2023 and 2024 and qualitative analysis of the information gathered about them. We wanted to analyze if this discourse was then reflected in the project's focus and activities.

In total, we conducted different queries to the Community Research and Development Information Service of the EU (CORDIS)[1] system, which stores the relevant information about research projects funded by the European Commission. For this search, we combined the keywords in the query: "Artificial Intelligence" AND "Law Enforcement Agencies" to identify the relevant projects in this topic funded under the H2020 framework program.[2] As mentioned

above, our focus is on developing AI technologies for law enforcement applications due to the risks and challenges that these technologies can pose to fundamental human rights and also because there is already existing analysis of the ethics compliance of EC-funded projects in other domains.[3]

We gathered information from 25 different projects related to 13 different work programs that got a total contribution from the EC of 101,457,752.12 €. The average contribution of the EC to these projects is more than 4.2 million per project in the 15 different research programs identified (see Table 1).

Most of the analyzed projects coincide with the policy development reflected in the chronology described in Fig. 1. This should result in the incorporation of these policy goals and principles in the implemented projects. The specific 25 projects we have analyzed are the following: (see Table 2).

As we can see in Table 2, most of the projects were funded in "Ensure privacy and freedom, including on the Internet and enhance the societal, legal and ethical understanding of all areas of security, risk and management" program. In our first analysis, we examined the information available in CORDIS for each project. A typical project section in CORDIS includes the following information: factsheet results in brief, reporting, and results. We analyzed all this information from the projects mentioned

---

Table 2 EU research projects involved in AI for law enforcement

Acronym	Start Date	End Date	Program Code	EC Max Contribution
SpeechXRays	2015-05-01	2019-04-30	DS-02-2014	4,102.467,00 €
RANGER	2016-05-01	2019-12-31	BES-01-2015	7,992.312,50 €
ASGARD	2016-09-01	2020-11-30	FCT-01-2015	11,992.553,25 €
DANTE	2016-09-01	2019-02-28	FCT-06-2015	4,998.527,88 €
iBorderCtrl	2016-09-01	2019-08-31	BES-05-2015	4,501.877,50 €
PROTECT	2016-09-01	2019-08-31	BES-06-2015	4,981.752,50 €
VIRT-EU	2017-01-01	2019-12-31	H2020-EU.2.1.1	1,999.951,25 €
FIRE-IN	2017-05-01	2022-10-31	H2020-EU.3.7.6	3,496,241.25 €
GHOST	2017-05-01	2020-04-30	H2020-EU.3.7.4	3,603,831.75 €
ILEAnet	2017-06-01	2022-05-31	H2020-EU.3.7.6	3,482,146.25 €
DAREnet	2017-09-01	2023-02-28	H2020-EU.3.7.6	3,500,000.00 €
eNOTICE	2017-09-01	2023-08-31	H2020-EU.3.7.6	3,497,735.00 €
I-LEAD	2017-09-01	2023-02-28	H2020-EU.3.7.6	3,483,716.00 €
MAGNETO	2018-05-01	2021-04-30	H2020-EU.3.7.6	5,320,475.00 €
SEREN 4	2018-05-01	2021-08-31	H2020-EU.3.7.6	1,999,953.75 €
COPKIT	2018-06-01	2021-09-30	H2020-EU.3.7.6	4,986,973.75 €
EXERTER	2018-06-01	2023-05-31	H2020-EU.3.7.6	3,498,868.75 €
MEDEA	2018-06-01	2023-05-31	H2020-EU.3.7.6	3,495,843.75 €
NO FEAR	2018-06-01	2023-05-31	H2020-EU.3.7.6	3,495,957.50 €
ARCSAR	2018-09-01	2024-02-29	H2020-EU.3.7.6	3,492,021.25 €
PEN-CP	2018–09-01	2024-01-31	H2020-EU.3.7.6	3,496,253.13 €
D-ID	2018-12-01	2019-05-31	H2020-EU.3	50,000.00 €
Humane AI	2019-03-01	2020-04-30	FETFLAG-01-2018	999,250.00 €
SHERPA	2019-10-01	2023-09-30	RUR-01-2018-2019	4,999,747.50 €
NoBIAS	2020-01-01	2023-12-31	H2020-EU.1.3	3,989,295.61 €

above to contrast how they followed the EU-defined principles and rules for a trustworthy right-based AI. Instead of doing the analysis manually, we automatized the text search. We manually collected the CORDIS web pages containing information about selected projects. Then, we automated the information gathering process and stored them in Excel files. Regarding web searches, we did a similar process: we automated the queries, searching only inside the domains of the projects (i.e., only searching on web pages that are in https://www.h2020-dante.eu/ and counting the number of pages that contain the word "gender"). We collected the total number of results obtained and stored the data in an Excel file.

To analyze previously defined aspects that an AI trustworthy system should include (be based on human rights and core European values; endorse the principle of non-discrimination and integrate a solid ethical component and societal impact), we first established our selection criteria based on a–c, selected the variables for each aspect, and established three categories representing different elements of trustworthy AI: foundation, consequences, and regulation (see Table 3). These categories and variables were defined inductively based on the analysis of the founding documents of the EU's policy objectives.

The final choice related to values, human rights, discrimination, gender, race, and ethics as these are related to the main biases and risks that these technologies pose according to the literature and to the EU's documents and guidelines (see Sect. 1: The European Digital Strategy; (Skeem and Lowenkamp [24]), and we checked whether they appeared in each of these projects' websites and related documents.

The selection of the words was made with special care. As can be seen, all selected words are well established not only in the literature on the subject, but also in the media and social networks. In our view, addressing some of the points such as the impact of artificial intelligence on human rights and ethics makes the use of these words unavoidable. However, languages are very rich in nuances, and it may be possible to address these issues without using the exact words.

We performed this analysis through a multiple-step process in which we analyzed the total number of results in the search engine for a website, then the number of pages where the analyzed word appeared, and then we calculated the total percentage of words appearing on the website. We conducted this analysis because we believe that if these projects incorporated these values and principles into their way of doing things and the design of their technology, this should be reflected in the documents, deliverables and

results disseminated. We conducted the same operation for each of the following values and then calculated the final result per project, as the graphs below show:

Subsequently, for a more in-depth analysis of the results, we selected 5 projects (ARCSAR, ASGARD, COPKIT, PROTECT, and VIRT-EU) and analyzed the content of their websites. The selection of this project was made based on the results of the projects in the previous phase.

We have chosen to analyze the PDF content hosted on the websites in this analysis. This content, mainly focused on dissemination materials and project deliverables, is a representative sample of the website's content. After extracting the text from the files, we displayed the information in a word cloud. We used the Python libraries *PyPDF2* for text extraction and *word cloud* for the creation of the word cloud. We have also used the *NLTK library* to tokenize the words and remove stopwords. Specifically, the following steps have been performed:

– Extra whitespace has been removed, using regular expressions.
– Punctuation marks have been removed, using regular expressions.
– The words have been normalized, converting each one to lowercase.

During the development of our research, we identified several challenges and shortcomings:

• We tested the tool and the methodology only in several projects. In order to improve and consolidate our findings, further development is needed to extend this analysis to all funded projects developing AI.
• Our findings relate only to the project's dissemination material (e.g., websites and deliverables). However, whether we can extrapolate impact measurement to the projects themselves is unclear. Due to the security constraints of this type of project, the provided information is limited, and many deliverables are labeled as "confidential."
• We realized that our methodology penalizes those projects that have generated more content. For the same number of words flagged, those projects with fewer pages show higher or better results than those with more pages. Also, the other way around, those projects with much content, maybe less represented. In the same vein, some projects are related to the topic but may be broader (for example, not only includes LEAs, but also other emergency services).

4 Results and discussion

Our text-based qualitative analysis led to some findings that we try to summarize in the following points.

4.1 The level of trustworthiness of AI for Law Enforcement EU-funded research projects is very low

Our analysis reveals that while the trustworthy-related concepts (values, human rights, discrimination, gender, race, and ethics) appear in the principal project analysis, they do not appear in the dissemination and deliverables (see Fig. 2). We aggregated the scores of our three categories (see Table 3), so the maximum available score is 300% (i.e., 100% per category). As seen in Fig. 2, few projects get a score higher than 50% (e.g., one of the analyzed terms appears on one page in two). Only one project (VIRT-EU) has some analyzed terms on over half of its web pages.

After analyzing all the scores, the most noticeable finding in our analysis is that most project scores are shallow, i.e., they barely account for any of these concepts in their websites, dissemination, and communication (see Fig. 3). Most of the projects achieve a score under 0.11 (11%) in all analyzed categories: 20 projects for the category "Foundation for Trustworthy" (human rights, values), 21 in Adverse consequences known to us (discrimination, race, gender), and 17 in bases as regulation (ethics):

Our selection of projects corresponded to various calls within the H2020 program. Specifically, the selected projects belonged to 13 different calls. As the Work program "Ensure privacy and freedom, including on the Internet and enhance the societal, legal and ethical understanding of all areas of security, risk and management,"[4] gathered most projects identified in our search (13) and it was also the Work Program that received the higher funding, and that is why we zoomed into it. When we did so, our analysis revealed mixed results for the levels of trustworthiness (see Fig. 4). Two projects score well, while the rest do not exceed 3.5%. These data may indicate that the definition and drafting of the calls for research projects are insufficient to obtain deliverables and dissemination materials that are fully aligned with the values of the European Commission. Therefore, the sensitivity of the consortia remains the key to integrating these values in the dissemination of project results.

From these graphs and our analysis, we can conclude that an analysis of the dissemination of information on research projects funded by the EU and geared toward developing AI solutions for law enforcement show a disconnect. Such disconnect appears first and foremost concerning their goals

---

[4] See https://cordis.europa.eu/programme/id/H2020-EU.3.7.6.

Fig. 2 Level of trustworthiness in EU research projects for AI in law enforcement

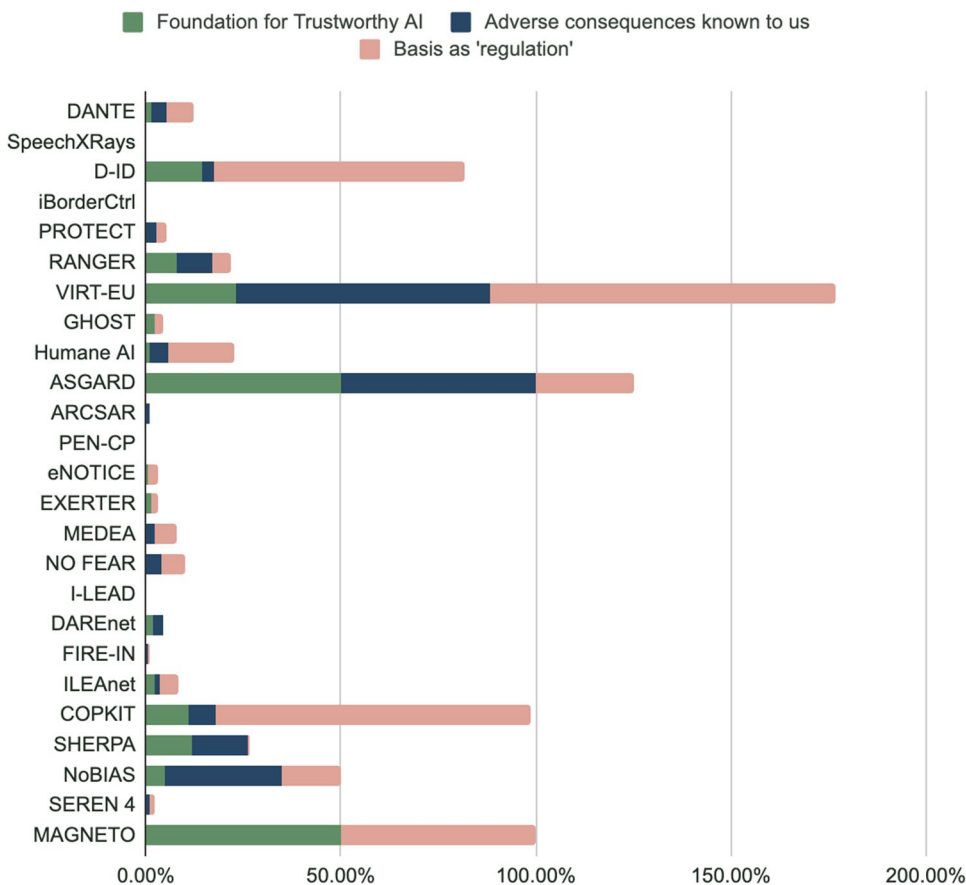Legend: ■ Foundation for Trustworthy AI ■ Adverse consequences known to us ■ Basis as 'regulation'

Table 3 Categories for the foundation of Trustworthy AI based on our analysis

Category	Variables
Foundation for Trustworthy AI	Values
	Human rights
Adverse consequences known to us	Discrimination
	Gender
	Race
Basis as 'regulation'	Ethics

and promises as stated on their website and the subsequent project execution, implementation, and technology development, which seems to vanish somehow. The second disconnect refers to the distance between the EU's policy goals supporting a trustworthy, responsible, non-discriminatory artificial intelligence in Europe and the actual funding poured toward developing such tools that do not consider much of the trustworthiness pillars that sustain such AI gears.

4.2 There is barely trustworthy AI Word Usage in EU AI for Law Enforcement research projects

As some authors point out, language plays a role in constructing reality [6, 29, 30]. Using words such as gender, race, discrimination, or human rights in this type of project's dissemination material may help make these concepts more prominent in this area. A clear example of the impact of language and the incorporation of certain buzzwords in the technological field can be seen in most of these institutions' corporate social responsibility policies.[5] On the contrary, not mentioning these words may represent a lack of consideration of essential concepts that should have a more prominent role in the project's foundations and execution.

Our analysis includes investigating much text within various websites and disseminating materials. The average number of pages we analyzed per project is 170, and the results are diverse. It seems clear that some terms are gaining weight, especially those related to ethics. Nevertheless,

---

[5] Companies such as Google, Microsoft and IBM have published ethical principles for the development of their AI technology. They incorporate these words to show a concern for the respect of these values and ethical principles and a concern to combat biases.

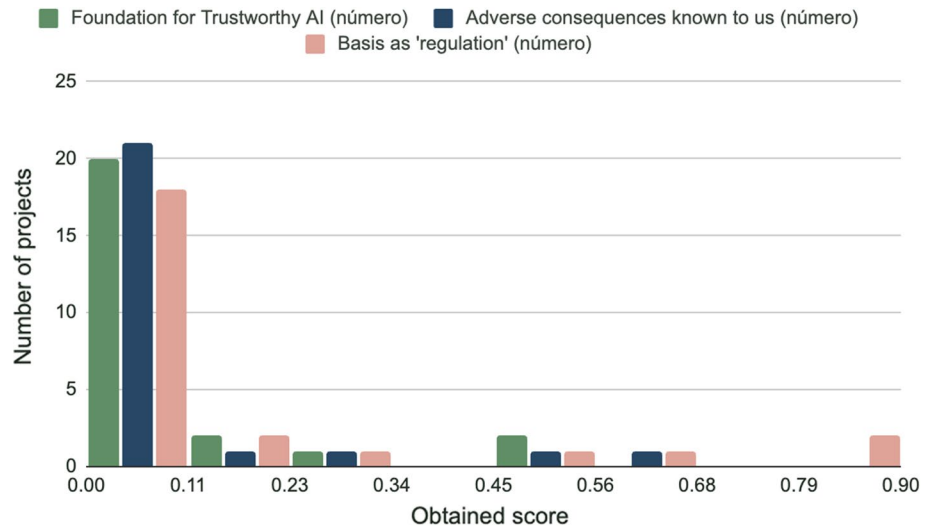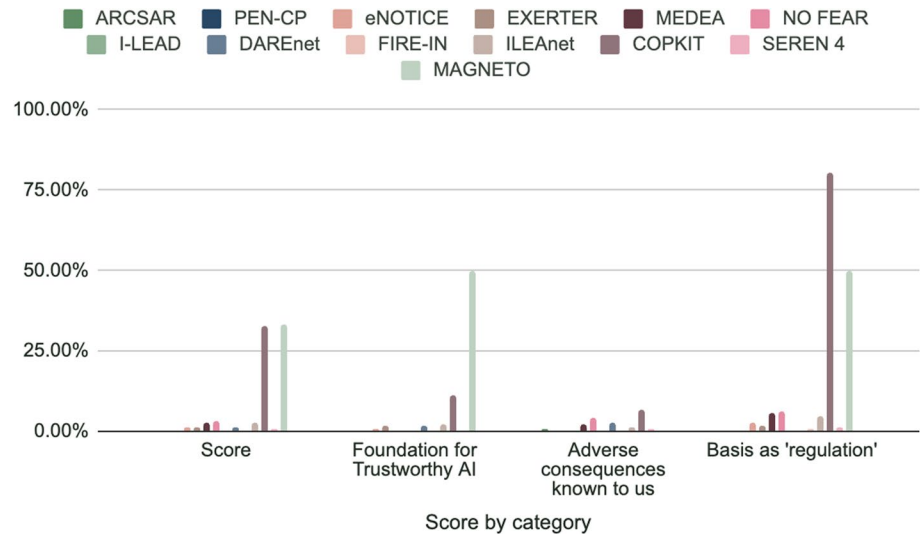**Fig. 3** Overall score of the level of trustworthiness

Fig. 4 Overall score of the level of trustworthiness: project overview

there is no significant progress in using these terms in the projects analyzed. The following graph also shows the uneven distribution of projects over time (for example, only two projects end in 2022, while nine end in 2022). This distribution may also partly affect the interpretation of the results (Fig. 5).

Although the increase in word usage may indicate an increase in attention paid to certain concepts, such interpretation needs to be asserted in a specific context [6]. In this case, the concepts of discrimination, race, gender, and values are considerably lower. A salient result is how little race and gender are featured on all of these websites throughout the years, given that the literature has consistently highlighted how some of these systems discriminate against more significant numbers of dark-skinned and female citizens (Buolamwini and Gebru [5]; Rademacher et al. [21]) [22].

By using and building upon the results of our findings, we stress the role of the EU as a key AI actor, as required by European AI Strategy (pillar 1), while at the same time ensuring that the technologies fulfill the ethical requirements defined by European values and the fundamental-rights framework. In this respect, more work must be done for the EU to ensure that policy goals and research programs align to respect fundamental rights in a context where the stakes are high, as in the case of AI for law enforcement.

To perform a deeper analysis of this point, we have analyzed the words in the deliverables of 5 projects (ARCSAR, ASGARD, COPKIT, PROTECT, and VIRT-EU) and extracted the content of the PDF files. After preprocessing, we generated a word cloud for each of the projects (Fig. 6).

As we can see, some of the limitations identified in the methodology section are reflected. The ASGARD project does not have any of the analyzed terms among its most
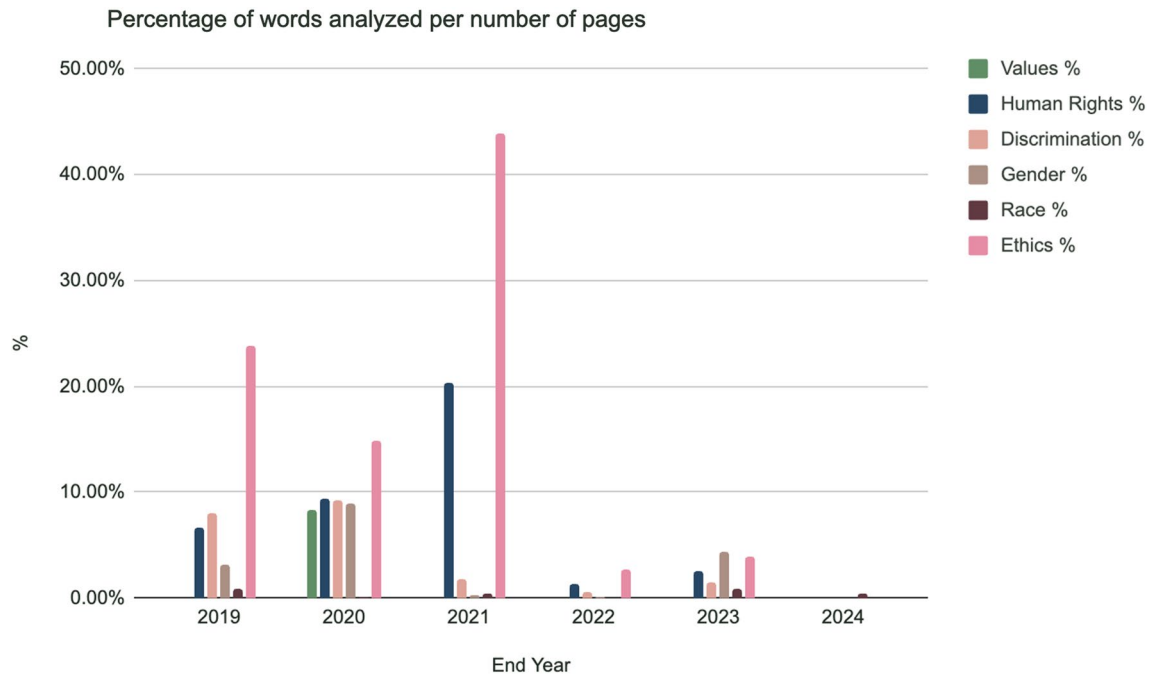
Fig. 5 Percentage of trustworthiness-related words analyzed in EU AI for Law Enforcement research projects

(a) ARCSAR project

(b) ASGARD project

(c) COPKIT project

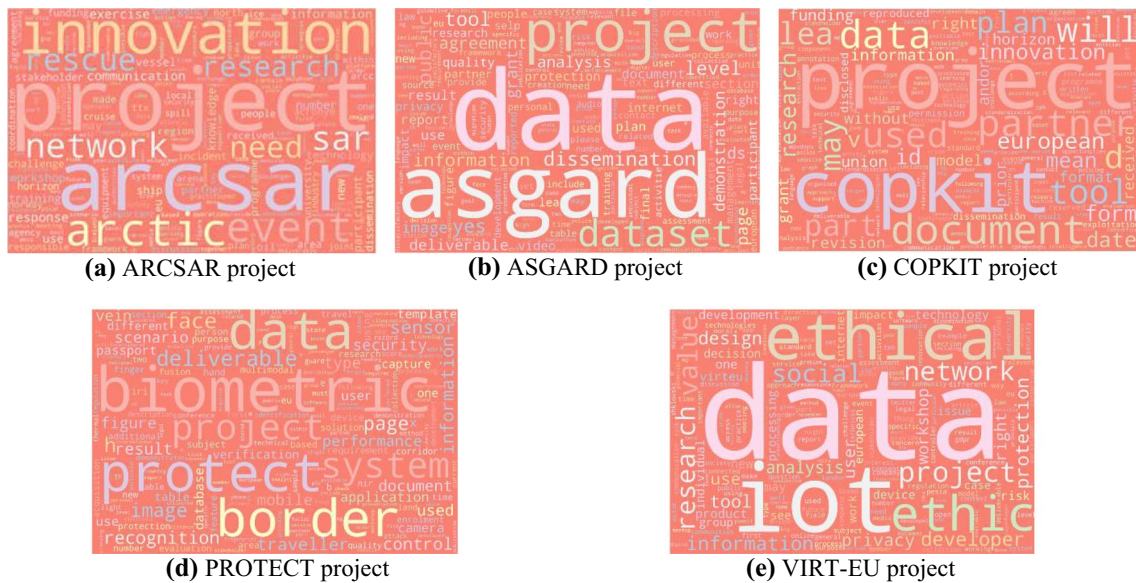**(d)** PROTECT project

(e) VIRT-EU project

Fig. 6 Wordclouds of analyzed projects

named words. However, given the few pages that make up its website, it obtains high results in our metric. We can also observe that the analyzed projects would obtain a worse result in general if we only analyzed the PDF files. In general, there are few appearances of the analyzed words in the analyzed deliverables and dissemination material.

4.3 The lack of citizen and NGO involvement in AI for law enforcement is apparent

A recurrent concern from the EU is that there is little or practically no citizenry involvement in research projects concerning law enforcement. Not involving citizens in such domains is salient because while AI can support the execution of law enforcement tasks, it can also have adverse

Fig. 7 Lack of NGOs in EU-funded AI projects for security and law enforcement
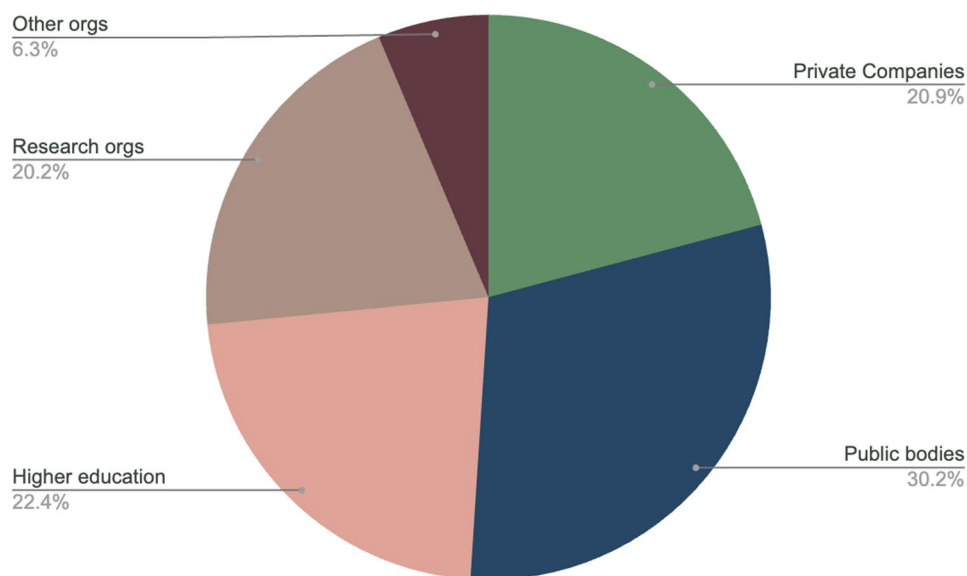
consequences on the fundamental rights of citizens. Aware of this issue, in 2020, the EU tried to change this narrative via different calls for proposals. For example, in the call for proposals "H2020-SU-AI-2020: Artificial Intelligence and security: providing a balanced assessment of opportunities and challenges for Law Enforcement in Europe," under the specific topic "SU-AI03 Human factors, and ethical, societal, legal and organizational aspects of using Artificial Intelligence in support of Law Enforcement,"[6] the EC stressed the importance of engaging "citizens and improving their perception of security." To this end, in this call, the EC encouraged specifically "consortia involving relevant security practitioners, civil society organizations as well as Social Sciences and Humanities experts."

This narrative shift seemed to be a big step toward one of the fundamental components of trustworthy AI. However, a close look at the call reveals something different. According to the information provided by the EC, this topic had considerably less available funding (1.5 M€) than the other two topics dedicated to AI development for law enforcement

(17 M€). Although the call gathered more proposals (a total of 13 big consortiums applied), the resource provision and availability for analyzing the ethical, legal, and societal consequences of such technology and the citizen involvement were considered more than 11 times less than the development of technology. Apart from further deepening existing gaps in research funding for social science research, such a move relates more closely to ethics washing than to a serious consideration of the issues at stake for EU society.

To understand whether this was an isolated case or whether this EU practice was a recurrent issue, we looked closer at the partner configuration of the analyzed projects. The first signal that there is a disconnect is the classification that the EC makes in CORDIS about the organizations: "Public bodies (excluding Research Organizations and Secondary or Higher Education Establishments)," "Higher or Secondary Education Establishments," "Higher or Secondary Education Establishments," "Private for-profit entities (excluding Higher or Secondary Education Establishments)" and "Other." The NGOs and other civic organizations are inside this "Other" category.

The second finding is that the results for this category show that they are clearly underrepresented (Fig. 7):

It is not easy to imagine that the project proposals are equipped with the necessary knowledge about these tools' impacts on society if the groups representing parts of society are not considered. Although universities usually play such a role, and the groups focusing on ethical and legal aspects of these technologies have become more prominent, it is nevertheless true that academia has its problems. To better understand the worries and concerns that certain groups have about society, they should be invited to participate in discussions revolving around technology development that may affect them directly or indirectly.

---

[6] See the call 'Human factors, and ethical, societal, legal and organizational aspects of using Artificial Intelligence in support of Law Enforcement' within the program Secure societies—Protecting freedom and security of Europe and its citizens, at: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/su-ai03-2020;callCode=H2020-SU-AI-2020;freeTextSearchKeyword=;matchWholeText=true;typeCodes=1;statusCodes=31094501,31094502,31094503;programmePeriod=null;programCcm2Id=null;programDivisionCode=null;focusAreaCode=null;destination=null;mission=null;geographicalZonesCode=null;programmeDivisionProspect=null;startDateLte=null;startDateGte=null;crossCuttingPriorityCode=null;cpvCode=null;performanceOfDelivery=null;sortQuery=sortStatus;orderBy=asc;onlyTenders=false;topicListKey=callTopicSearchTableState.

In our experience, the invitation to participate in civil society is a necessary but not sufficient condition to achieve the goal requested by the European commission. This invitation has to be done in a space where both actors are trusted in the intentions and in the capabilities of counterparty. To create this space, it is necessary to create small actions to build this confidence, particularly in topics where the tensions between actors are strong (for example, NGOs and LEAs in human trafficking topic). The preparation stage of one proposal could be an interesting starting point to build this confidence.

5 Conclusions and future work

There is a growing interest in making AI trustworthy, AI that respects and supports fundamental rights, at least in Europe. However, our research findings reveal existing disparities between such good intentions enshrined in the EU policy goals for trustworthy AI and the actual EU research practices. This disconnect appears on different levels, some more noticeable than others:

– Research projects for AI in law enforcement do not echo the trustworthiness levels put forward by the EU institutions;
– The language used in the project proposal and dissemination activities shows disparity and an overall disregard for trustworthy elements;
– In some projects, the funding made available for ethical and legal aspects is considerably lower than the budget for the development of the technology;
– The project partner composition shows a lack of involvement of NGO and citizen organizations.

In a context like law enforcement, all this may have adverse consequences for society. Future work will thoroughly review existing literature and online resources to investigate the landscape of AI for law enforcement, identifying market niches, underexplored applications, and their associated societal challenges and ethical barriers.

Equipping the ecosystem surrounding these technologies with a multidisciplinary understanding of the multiple repercussions of AI for law enforcement is essential to remove barriers that prevent the uptake and acceptance of these technological advancements in security. For now, research projects dealing with law enforcement should find ways to facilitate NGO and citizen involvement and equip themselves with robust interdisciplinary knowledge to help them anticipate potential pitfalls in such a delicate arena. At the same time, more oversight from the EC should be put in place to understand how the EU policy goals align with the research agenda of the Union.

Declarations

References

1. Almeida, D., Shmarko, K., Lomas, E.: The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. AI and Ethics **2**, 377–387 (2022)
2. AlZoubi, O., Hussain, M. S., & Calvo, R. A.: Affect-aware assistive technologies. In: Assistive technology for cognition. Psychology Press, pp 61–78 (2014)
3. Apostolakis, K.C., Dimitriou, N., Margetis, G., Ntoa, S., Tzovaras, D., Stephanidis, C.: DARLENE–Improving situational awareness of European law enforcement agents through a combination of augmented reality and artificial intelligence solutions. Open Res. Europe **1**(87), 87 (2022)
4. Aroyo, A.M., De Bruyne, J., Dheu, O., Fosch-Villaronga, E., Gudkov, A., Hoch, H., et al.: Overtrusting robots: Setting a research agenda to mitigate overtrust in automation. Paladyn J. Behav. Robot. **12**(1), 423–436 (2021)
5. Buolamwini, J., Gebru, T.: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on fairness, accountability and transparency. PMLR, 2018
6. Cañete-Valdeón, J.M., Galán, F.J., Toro, M.: The intentional relationship of representation between the constructs of a language and reality. Data Knowl. Eng. **68**(2), 173–191 (2009)
7. COM 2019.: Building Trust in Human Centric Artificial Intelligence (2019)
8. COM 2021 (205 final).: Communication on Fostering a European approach to Artificial Intelligence (2021)
9. Dechesne, F., Dignum, V., Zardiashvili, L., Bieger, J.: Long-term Research Strategy for Artificial Intelligence and Ethics at the Police. Long-term Research Strategy for Artificial Intelligence and Ethics at the Police (2019). https://scholarlypublications.universiteitleiden.nl/access/item%3A2983930/view. Accessed 3 Nov 2022
10. European Commission: On artificial intelligence—a European approach to excellence and trust. European Commission (2019). https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Accessed 3 Nov 2022
11. Fosch-Villaronga, E. Poulsen, A. (2022) Diversity and inclusion in artificial intelligence. In: Custers, B., Fosch-Villaronga, E (eds)

Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice. Information Technology and Law Series, vol 35. T.M.C. Asser Press, The Hague, pp 109–134. https://doi.org/10.1007/978-94-6265-523-2_6

12. Floridi, L.: The European legislation on AI: a brief analysis of its philosophical approach. Philos Technol **34**(2), 215–222 (2021)

13. Fuster, G.G.: Artificial Intelligence and Law Enforcement. Impact on Fundamental Rights. Policy Department for Citizen's Rights and Constitutional Affairs (2020). https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf. Accessed 3 Nov 2022

14. Hardyns, W., Rummens, A.: Predictive policing as a new tool for law enforcement? Recent developments and challenges. Eur. J. Crim. Policy Res. **24**(3), 201–218 (2018)

15. HLEG.: Pilot the Assessment List of the Ethics Guidelines for Trustworthy AI. (2019). https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal

16. High-Level Expert Group on AI: Ethical Guidelines for Trusworthy AI. European Commission (2019). https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed 3 Nov 2022

17. Noble, S. U.: Algorithms of oppression: How search engines reinforce racism. New York University Press, 2018

18. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts Com/2021/206 Final

19. Sætra, H.S., Fosch-Villaronga, E.: Research in AI has implications for society: how do we respond? The balance of power between science, ethics, and politics. Morals Mach. **1**(1), 62–75 (2021). https://doi.org/10.5771/2747-5182-2021-1-62

20. Raaijmakers, S.: Artificial intelligence for law enforcement: challenges and opportunities. IEEE Secur. Priv. **17**(5), 74–77 (2019)

21. Rademacher, T.: Artificial intelligence and law enforcement. In: Wischmeyer, T., Rademacher, T. (eds.) Regulating Artificial Intelligence, pp. 225–254. Springer, Cham. (2020)

22. Ritchie, A.J., Jones-Brown, D.: Policing race, gender, and sex: a review of law enforcement policies. Women Crim. Just. **27**(1), 21–50 (2017)

23. Rosner, L.: The technological fix: how people use technology to create and solve problems. Routledge, New York (2013)

24. Skeem, J.L., Lowenkamp, C.T.: Risk, race, and recidivism: predictive bias and disparate impact. Criminology **54**(4), 680–712 (2016)

25. Sousa, W.G., de Melo, E.R.P., Bermejo, P.H.D.S., Farias, R.A.S., Gomes, A.O.: How and where is artificial intelligence in the public sector going? A literature review and research agenda. Gov. Inf. Q. **36**(4), 101392 (2019)

26. Ulnicane, I.: Artificial intelligence in the European union: policy, ethics and regulation. In: Hoerber, T., Weber, G., Cabras, I. (eds.) The Routledge handbook of European integrations. Routledge, New York (2022)

27. Veale, M., Zuiderveen Borgesius, F.: Demystifying the draft EU artificial intelligence act-analysing the good, the bad, and the unclear elements of the proposed approach. Comput Law Rev Int **4**(22), 97–112 (2021)

28. White Paper on AI 2020.: COM(2020) 65 final, Brussels, 19.2.202

29. Wieder, D.L.: Language and social reality. De Gruyter Mouton, Berlin (2015)

30. Willig, C.: Discourse analysis. Qual Psychol Pract Guide Res Methods **2**, 160–186 (2003)

31. Završnik, A.: Criminal justice, artificial intelligence systems, and human rights. ERA Forum **20**(4), 567–583 (2020)