Voices from the Frontline: Revealing the AI Practitioners' viewpoint on the European AI Act

Fiona Koh
Liechtenstein Business School
University of Liechtenstein
fiona.ruettimann@uni.li

Kathrin Grosse
VITA Lab.
École Polytechnique Fédérale de Lausanne
kathrin.grosse@epfl.ch

Giovanni Apruzzese
Liechtenstein Business School
University of Liechtenstein
giovanni.apruzzese@uni.li

Abstract

Artificial intelligence (AI) is increasingly used in an ever larger number of industries. Alongside this development, however, abundant works argue that AI-driven systems are lacking in terms of safety, ethics and transparency. As a direct consequence, the European Commission is working on the AI Act—a regulation designed to ensure a trustworthy development of AI that is respectful of its end-users' well-being. Despite the impact this law will have on the AI industry, few studies cover more than two or three aspects of the AI Act from the industry's perspective. In this paper, we attempt to close this gap and interview 21 companies to understand their holistic view on the upcoming regulatory landscape. We find that while the overall opinion on the AI Act is positive, there is a need for further resources like personnel and information to increase legitimacy. We further shed light on our companies' desiderata for the AI Act: more fine-grained regulation, and more AI expert input. Lastly, we identify avenues for future research, entailing machine unlearning and a deeper understanding of industry's perception on current legislation.

Keywords: Artificial Intelligence, European AI Act, GDPR, Security and Privacy, Regulation and Legislation

1. Introduction

The worldwide interest towards artificial intelligence (AI) is constantly growing. This technological paradigm has now passed the path to maturity, and many operational information systems are now powered by AI (Jiang et al., 2022). In a sense, almost no day goes by without hearing about AI. However, while many stories narrate positive AI developments (Haefner et al., 2021), others portray AI with shades of gray (Petersson et al., 2022), and some even claim that AI can be a threat to organizations (Mirsky et al., 2022) and society (Caldwell et al., 2020). The stark reality is that the field of AI is advancing at such a rapid pace that even its creators cannot envision where these technologies can truly lead (Dwivedi et al., 2023).

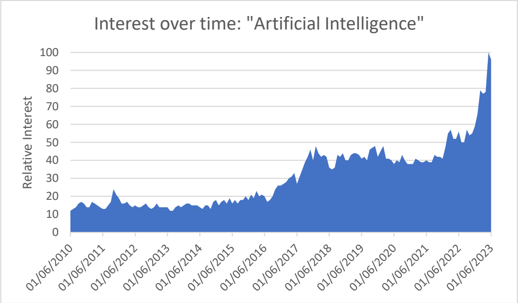

Figure 1. Worldwide search interest of "artificial intelligence" since 2010 (source: Google Trends).

In an effort to prevent the (unforeseeable) risks and harms that originate from real-world deployments of AI, many experts advocate the necessity of AI-specific **regulations** (Stone et al., 2016). Such pleas were heard by sovereign entities and resonated well in Europe: After taking a clear stance (in 2020) on the broad topic of AI – whose development should be guided by "trustworthiness" – the European Commission devised the AI Act in 2021 (refer to §2 for background). The official enactment of the AI Act is expected to significantly affect the AI industry in Europe (Canhoto & Clear, 2020). In particular, *companies* having AI as their core business will have to comply with the rules discussed therein—some of which are very restrictive for sensitive applications of AI.

Many scientific works have analysed the AI Act itself (such as Kaminski (2021)), and some have investigated the viewpoint of practitioners on some points contemplated in its regulations (e.g., AI ethics (V. Johansson et al., 2022)). However, we observed that no study elucidated the viewpoint of AI practitioners with respect to the whole spectrum of themes envisioned in the AI Act (§2.3). This absence is concerning, given the impact that the AI Act will have on the future of AI (Gragousian, 2022).

We aim to close this gap by investigating the position of AI companies in light of the recent European regulatory landscape. We carry out semi-structured interviews with 20 technical employees affiliated with 21 AI companies located in Europe (§3). Our questions entail topics such as AI security, ethics, performance assessment, and data governance. To provide holistic coverage, we also focus on the interplay between the (upcoming) AI Act and the (existing) GDPR. We present our findings according to existing legislation (§4.1); next, we focus on AI security (§4.2), and then report the perception of the AI Act (§4.3). Motivated by our findings, we also carry out a small survey which reveals the *general population*'s sentiment towards the use of the term "AI" by companies (§4.4) Finally, we coalesce our results and derive actionable implications (§5.2).

To summarize, our work provides a **threefold contribution** to the field of AI in governance: **(1)** *We elucidate the relationship between AI practitioners and ongoing legislation in Europe.* While many companies seemed prepared for the AI Act, more resources are required to ensure its compliance, and there is a lack of experts that can help address legal issues. **(2)** *We link industry and political agenda by reporting desiderata for the AI Act according to our interviews.* While most participants welcomed regulations and perceived them as an opportunity, they also wished for more fine-grained regulation—for which more input from AI experts is desired. **(3)** *We identify future avenues of academic research.* This entails more work on machine unlearning and real-world AI security research.

2. Background and Related Work

We first summarize the most recent advances of AI (§2.1) and introduce the European AI Act (§2.2). Then, we describe the research gap addressed by our paper: the lack of studies focusing on the perspective of AI companies in the upcoming regulatory landscape (§2.3).

2.1. Current state of AI

AI is **ubiquitous** today (Jiang et al., 2022). After nearly a decade during which machine and deep learning became increasingly popular in research (Jordan & Mitchell, 2015; LeCun et al., 2015), AI-driven systems have become pervasive also in practice. Various domains now leverage AI, such as healthcare (Bohr & Memarzadeh, 2020), cybersecurity (Kshetri, 2021), finance (Königstorfer & Thalmann, 2020), autonomous driving (Ning et al., 2021), management (Haefner et al., 2021), social media (Laacke et al., 2021), and even education (Ouyang et al., 2022). In particular, the advent of ChatGPT was recognized by many as one of the most disruptive advancements in IT (Dwivedi et al., 2023), given that it provided enormous potential accessible by everyone—everywhere and everytime. As a matter of fact, many industrial sectors now *depend* on ChatGPT (George & George, 2023) and AI in general, with many companies having AI as their core-business. For instance, the AI market share amounts to $200B in 2023, and it is expected to grow to $2 *trillion* by 2030.[1]

Unfortunately, amidst the explosive interest in AI (see Figure 1) within our society, the careless adoption of AI leads to many **risks**. Indeed – like any man-made product – AI is not perfect, and its *vulnerabilities* can be exploited by attackers, who can evade, poison (Biggio & Roli, 2018) or even *steal* an AI model (Tramèr et al., 2016). Alternatively, *improperly trained* AI can lead to inadvertent discrimination of humans (Wei & Zhou, 2023) and/or increased psychological risk (Dwivedi et al., 2023). Finally, AI can also be *abused* to, e.g., spread false news[2] (Reisach, 2021), or as a weapon to carry out cyberattacks—leading to security (Mirsky et al., 2022) or privacy (Tricomi et al., 2023) violations.

In light of these risks, many experts suggest **exercising caution** in the indiscriminate deployment of AI (Brundage et al., 2018; Stone et al., 2016). For instance, Hoffmann-Riem (2020) argue that AI companies are unlikely to always comply with ethical standards, and endorse the enactment of legal rules that bind the developers of AI systems and their end-users. A similar opinion is given by Boddington (2017), who point out that AI systems suffer from poor transparency and accountability. In response to these "calls for regulations", the European Commission (EC) released a white paper in 2020 (European Commission, 2020b), describing how the EC aims to promote a fair and risk-free usage of AI within the European boundaries. Accordingly, the first step to reach such an ambitious goal is the enactment of the so-called "AI Act".

2.2. The European AI Act

The AI Act is a legislation proposed by the EC, which aims to regulate AI in Europe (European Commission, 2021). Its first version was published in April 2021, and it was the first attempt at AI regulation

[1] Source: Statista (https://www.statista.com/statistics/1365145)

[2] Surprisingly, using AI to *counter* fake news is not very effective (Stachofsky et al., 2023).

proposed by a sovereign entity (Ruschemeier, 2023).[3] The AI Act reflects horizontal regulation and is rooted in the development of AI in compliance with safety and fundamental rights while promoting innovation and competitiveness by AI industries. Approved by the European Parliament on June 14th 2023, the AI Act is expected to be adopted in 2024, and its obligations will take effect within the next three years.

The AI Act follows a **risk-based approach**: It defines three different levels of risks: unacceptable[4] (which are prohibited in the EU), high[5] (which are highly regulated, e.g., by strict security measures), and low/minimal[6] (which are minimally regulated) risk. Regardless of the risk level, all AI systems are subject to requirements pertaining to transparency, explainability, accountability, as well as data-governance. Importantly, the AI Act is meant to complement (Campion et al., 2022) the already existing **General Data Protection Regulation** (GDPR). Indeed, AI systems rely on data-driven techniques, and hence may intrinsically raise privacy concerns, which must be accounted for.

Despite the potential benefits that the AI Act can bring to AI in Europe, it was not exempt from **critiques**. For instance, Ruschemeier (2023) points out an overall lack of clarity—such as a lack of a clear definition of "AI system", leading to uncertainty about the applicability boundaries of the AI Act itself. Furthermore, as remarked by Carter et al. (2020), technological progress advances at a much faster pace than regulations. Hence, there is skepticism on whether the AI Act will truly promote – and not hinder – the development of AI systems. For example, the release of ChatGPT opened up scenarios that were unforeseeable few months prior (Dwivedi et al., 2023). Indeed, the current European regulatory landscape (i.e., the GDPR and the upcoming AI Act) seeks to prioritize the well-being of end users. However, within this context, there is a specific group that is at risk of being left behind: companies whose core business relies on AI.

2.3. The (unknown) position of AI companies

According to Gragousian (2022), complying with European regulations while generating profits will be tough for AI companies. It is estimated that, by 2025, complying with the AI Act will cost European AI businesses nearly €30B (Mueller, 2021). Such figures can be prohibitive for small enterprises (e.g., startups): even the EC estimates that a single high-risk AI product could cost up to €400k. This is not surprising: for instance, if a user requests their data to be deleted, the corresponding company must manually delete the data-point from the database (for the GDPR), and then have the AI "forget" the data-point (for the AI Act) while ensuring that the performance is not excessively affected—all of which being operations that require extensive human effort (and which are open research problems (Bourtoule et al., 2021)). Furthermore, the EC released the "Assessment List for Trustworthy AI" (ALTAI), providing guidelines to develop more secure AI systems (European Commission, 2020a); however, ensuring that an AI system complies with such standards is costly, since it requires continuous assessments and manual fine-tuning (Apruzzese et al., 2023). As a result, investors may be reluctant to fund high-risk AI startups in the EU (Mundell, 2023).

Research Gap. Abundant research has investigated the perspective of AI companies w.r.t. ethics, security, or generic regulations of AI. However, to the best of our knowledge, no work *simultaneously* elucidated how AI companies *(i)* located in the EU are positioned in the *(ii)* current and future regulatory landscape, as well as their readiness level on *(iii)* ethics and security of AI.

Taken individually, some works have covered each of these topics. E.g., for regulations, Kaminski (2021) provide a critical perspective on the proposed EU AI Act, suggesting it falls short of its aim to ensure trustworthy AI; whereas Wachter et al. (2017) focus solely on the GDPR. For ethical AI development, Hagendorff (2020) highlights the limitations of prior ethical guidelines. Finally, Biggio and Roli (2018) extensively discuss the major security vulnerabilities of AI. Unfortunately, none of these papers provide the perspective of practitioners on the corresponding topic.

To the best of our knowledge, while there are papers that have dealt with regulation, the AI Act, performance measurements, explainability, bias mitigation or security – individually or in combination – no previous work has investigated *all* of these topics jointly by the means of *interviews*. We summarize all these related works in Table 1, where we also show the year of the study and the number of interviewees. The former is crucial, as older findings in a versatile area like regulation may not hold anymore. The only related study on the AI Act by Liebl and Klein (2022) is furthermore from an industrial initiative, not a peer-reviewed research paper.

We seek to **overcome these limitations** and provide a holistic perspective of European AI companies' viewpoints on the upcoming future of AI. Our study is important given that AI companies are responsible for putting innovation into practice, and hence help advance AI. Therefore, understanding their readiness level is crucial to ensure a smooth development of AI.

[3] Other countries, such as the USA, followed the EU steps much later (White House Office of Science and Tech. Policy, 2022).

[4] E.g., an AI that may nudge a child to overeat and become obese.

[5] E.g., an AI system deciding about access to a university.

[6] E.g., an AI that recommends which movies to watch.

Table 1. Related interview (and survey (*)) studies on regulation, the AI Act, performance measures, explainability and transparency, bias mitigation and cyber or AI security.

Author & Year	Pop. Size	Gen. Reg.	AI Act	Perf. Meas.	Expl. & Trsp.	Bias Mitig.	Sec.
Pumplun et al. (2019)	14	✓					
Rothenberger et al. (2019)	8	✓					
Rakova et al. (2021)	26			✓	✓	✓	
Jöhnk et al. (2021)	25			✓			✓
Petersson et al. (2022)	26	✓					
Liebl and Klein (2022)*	113	✓	✓				
V. Johansson et al. (2022)	12	✓			✓		
Hinsen et al. (2022)	25				✓	✓	
Bieringer et al. (2022)	15						✓
Meyer and Apruzzese (2022)	18	✓					✓
Leewis and Smit (2023)	42				✓		
Grosse et al. (2023)*	139						✓
Mink et al. (2023)	21						✓
Our work (2023)	21	✓	✓	✓	✓	✓	✓

3. Research Goal, Method, and Sample

This paper revolves around the following research question (RQ): "*how well positioned are AI companies within the upcoming European regulatory landscape?*" To investigate such an RQ, we carry out semi-structured interviews with technical employees of companies located within Europe and which have AI as their core business, asking questions pertaining to the themes related to the AI Act (e.g., GDPR, ethics, and security).

3.1. Methodology

Due to page limitations, we will only discuss the essential information. However, for transparency, we provide additional details (i.e., the questionnaire and the complete results of our study) in a public repository.[7]

Study and questionnaire design. While we had clear questions in mind to tackle our RQ, we also wanted to give our interviewees the chance to speak up and thereby share information we were not aware of (e.g., enable an *exploratory* study). To this end, we designed a questionnaire that would be presented to our participants in form of a *semi-structured* interview.[8] This allowed us to document unexpected replies and other comments. A similar methodology has been used in previous studies (Meyer & Apruzzese, 2022). After various meetings among the authors, we derived a set of 34 questions (q), divided into six groups. Specifically: (1) demographics, with 4q; (2) data storage, with 5q; (3) AI development, with 6q; (4) AI security, with 3q; (5) AI regulations, with 9q; (6) AI ethics, with 6q. Overall, these questions provide a holistic coverage of the themes tackled by this paper—all pertaining to our main research question. Once the questionnaire was finished, we conducted a pilot interview, from which we got only minor feedback which we incorporated.

Target Group and Recruiting. Our study is directed to any company that could potentially be affected by the AI Act. This entails all companies in Europe (including, e.g., Switzerland[9]) which use AI. To find companies, we browsed popular repositories (such as Traxcn or Clarafinds) as well as university incubators, looking for *active* companies that fell within our scope. We then (beginning in Oct. 2022) reached out to any potential candidate company – either via the contact information provided on their website, or by directly contacting some representative personnel – inquiring whether they were willing to participate in our research. To *avoid priming*, we limited our communication with these companies: we told them that the remote interviews were going to last ≈45m and a rough list of topics to be discussed within the meetings. This gave companies the chance to let us know if any topics within the interview would be off-limits. We did not offer any form of compensation to those who agreed. Furthermore, some companies only agreed to participate in the interviews under NDA. We are thus limited in the information we can disclose about our participants.

Interviews and data analysis. All interviews were conducted remotely by the first author of this paper. The first interview took place in Dec. 2022, for which a second author joined for support but without interfering. During the interview, participants were first shown a slide with a specific question (out of 34), which was read out loud by the interviewer. Afterwards, we revealed the possible answers (in case of closed-ended questions), which were also read out loud. The interviewer then manually registered the replies and possible comments on the question at hand. We gave the opportunity to interviewees to ask us for clarifications on specific questions: to *minimize bias*, we provided consistent answers (e.g., by revealing the definition of a term). To protect the privacy of our interviewees, we did not record the interviews. The last interview occurred in March 2023, resulting in a total duration of ≈3 months for our interviews (and ≈6 months for this entire study). We analysed the collected data by inspecting the replies to the questions and some extra remarks. All authors participated in these discussions, thereby enabling a *consistent interpretation* of our data.

3.2. Sample and limitations

In total, we interviewed 20 AI practitioners, each working at a different AI company; however,

[7] https://github.com/hihey54/hicss57-AIAct

[8] We treated our participants **ethically** (following the Menlo report); and the University of Liechtenstein approved our research.

[9] Even AI companies in Switzerland will be affected by the AI Act (https://www.pwc.ch/en/insights/regulation/ai-act-demystified.html)

one interviewee was currently employed by two AI companies. Hence, we carried out a total of 21 semi-structured interviews—each reflecting the position of a specific AI company. In our sample, 18 participants identified as male, 2 as female. The average age of the interviewees was 33 years, which is similar to studies within this population (Grosse et al., 2023). In terms of location, eleven (47.6%) companies were from Switzerland, five (23.8%) from Italy, and one (4.8%) each from Czech Republic, Austria, Ireland, and the Netherlands. Our sample covers six industries: business intelligence (ten, 47.6%), cybersecurity (four, 19%), healthcare and life sciences (four, 19%), mobility (one, 4.8%), education (one, 4.8%), and space (one, 4.8%). Finally, seventeen (80.1%) companies have less than 50 employees, one company more than 1000 employees, and the remaining three companies in between.

Limitations. We acknowledge that this study is based on a small and heterogeneous sample, and hence does not generalize. Our research is *exploratory in nature*: analyzing a similarly sized (e.g., (Hinsen et al., 2022; Jöhnk et al., 2021)) or even smaller samples (e.g., (Bieringer et al., 2022; Pumplun et al., 2019; V. Johansson et al., 2022)) is not unusual for exploratory studies. Our sample is biased towards males—but a similar skewed distribution affects also related studies (Bieringer et al., 2022). Finally, our sample only includes companies located in Europe: this is because our focus is on the AI Act—given that it is the most "mature" set of AI-specific regulations.

4. Results: what do AI practitioners say?

We present our main findings across three dimensions: compliance with existing regulation (§4.1), focusing on data governance and legal issues; AI security (§4.2), focusing on adopted practices and IP protection; and the AI Act itself (§4.3), focusing on ethics and readiness. To align the presentation with our questionnaire, we provide the reference to the specific question asked in our interviews (e.g., *Q.2.1* refers to the first question of part 2). We also showcase an additional online survey among the general population (§4.4). We provide the full results of our research in our repository.

4.1. Viewpoint on existing regulation (GDPR)

To be able to understand future regulation's effect on the companies, we first investigated their relationship towards existing, similar regulations. The GDPR was found to serve as a scaffold for perception (Bieringer et al., 2022) and is also European legislation, and consequently a good subject. We first inquired how companies dealt with two common issues related to GDPR: data storage location (for sensitive data, this has to be Europe) and data deletion (should be possible upon request). We then inquired whether the companies had already encountered legal issues.

Data storage location. We asked (*Q.2.1*) the companies where their data was stored. Seventeen interviewees (81%) stored their data in Europe, one interviewee (4.8%) both in- and outside Europe, and three interviewees (14.3%) outside Europe. In the latter cases, where data was hosted abroad, participants argued that client data was stored within Switzerland, that the company also operated outside Europe, or that the data stored abroad was not confidential. We consecutively asked (*Q.2.2*) these latter three companies how easily they could relocate their data—using a scale from 1 ("*very easy*") to 10 ("*very difficult*"). Two asserted that relocation would be rather easy (2 and 4); for the third, relocation would be difficult (8).

Data deletion. We asked (*Q.5.4*) the companies whether they could remove a single data-point upon request. Fifteen interviewees (71.4%) said they had such fine-grained data access. Five interviewees (23.8%) denied this, and one interviewee (4.8%) said they were not able yet, but were working on an implementation. As a follow-up question (*Q.5.5*), we inquired if the companies were able to delete many samples on a daily basis—using a scale from 1 ("*not equipped yet*") to 10 ("*very well equipped*"). Figure 2 illustrates the collected replies. As one company had said they were not able to yet, there is a total of 20 companies for this question. Five companies (25%) stated they were currently not equipped at all to handle many requests. Another five (25%) chose a low number (<5), indicating they were not well equipped. Seven (35%) felt that they were rather well equipped (> 4, < 10). Finally, three companies (15%) reported being very well equipped. (One company stated that data deletion does not apply to them, thus there are 20 data-points in Figure 2).

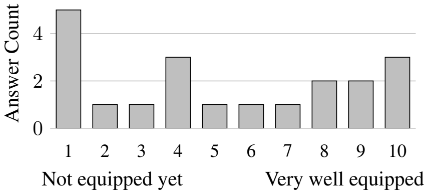

Figure 2. Replies to *Q.5.5* about ability to remove many samples daily upon request.

Legal issues. We asked (*Q.5.1*) whether and how often the companies had already faced legal issues related to their AI. One interviewee stated that their product was not on the market yet, leaving 20 companies. Seven companies had never faced legal issues, ten (50%) sometimes, and three (15%) said they often faced such issues. These latter three interviewees mentioned ignorance of regulations, unclear regulation, or legal disagreement between parties as reasons. We further inquired (*Q.5.3*) whether companies currently interact with an expert on legal topics. Seven (33%) did, three (14%) would like to, but did not know whom to contact. Another four (19%) were not in touch, but would like to do so soon. Seven (33%) were not in touch with a legal counselor and did not wish to.

Summary. Companies complied with basic requirements concerning data storage location and data deletion, but they also struggled with practical details, as visible in both data relocation and deletion. One reason was that current approaches did not scale, as described for data deletion. Finally, many companies were already facing legal issues, and some struggled to find the desired legal assistance.

4.2. Viewpoint on AI Security

Most attacks affect the AI's performance, leading us to inquire about performance assessments. Since previous works reported the uncertainty of practitioners towards AI security and their need for guidelines (Grosse et al., 2023), we first collected our companies' concerns on IP theft. We then conclude the section by focusing on recognized guidelines investigating the viewpoint on ALTAI (see §2.3).

Performance assessments. A common way to uncover attacks is by periodic system checkups. We thus asked the companies (*Q.3.1*) how often they measured the performance of their AI. Three (15%) interviewees did so daily, and nine (42.9%) once a week. Another three (15%) measured the performance monthly, two (9.5%) every three months, two (9.5%) once in six months, and finally one (4.8%) company reported that they never quantified the performance of their AI system. As a follow-up, we asked the interviewees how they measured that the AI's performance was maintained over time (*Q.3.2*). Twelve (57.1%) reported that an alert would trigger if performance was too low, and another twelve (57.1%) implemented retraining. Some companies also rely on human oversight (four, 19%), their clients (two, 9.5%), or new data, surveillance, or new models (one each, 4.8%). [10]

Intellectual property. We asked (*Q.4.2*) the companies whether they were aware that their IP may be at risk, e.g., that their AI model could be stolen. Eighteen (85.7%) were aware, while three (14.3%) were not. One interviewee commented that "*to have the AI system is one thing, but the data is what makes ours good,*" showcasing the importance of data. Further, we asked the companies (*Q.4.3*) whether their AI's IP was well protected—using a scale from 1 ("*not protected at all*") to 10 ("*extremely protected*"). We display these results in Figure 3. Interestingly, only four (19%) companies answered with a 5 or less.

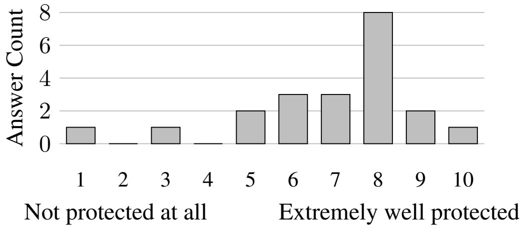

Figure 3. Replies to *Q.4.3* rating how well protected their AI in terms of intellectual property is.

AI security guidelines. We asked the companies (*Q.6.4*) whether they had heard of ALTAI: eighteen (85.7%) had never heard about ALTAI, while only two (9.5%) knew about it. We then inquired (*Q.6.5*) about the 7 individual elements of ALTAI. We plot the distribution over each element in Figure 4 (the caption states all elements of ALTAI). Intriguingly, albeit the majority of our interviewees do not know ALTAI, they implemented on average 4.8 of the 7 elements. Almost all (20, 95.2%) implemented technical robustness and safety, seventeen (80.9%) privacy, sixteen (76.2%) human oversight, fourteen (66.6%) environmental and social well-being, thirteen (61.2%) accountability, twelve (57.1%) transparency, and only eight (38.1%) diversity. We further asked (*Q.6.6*) whether this checklist would be helpful in the future for them. Sixteen (76.2%) answered positively, whereas three (14.3%) were not sure and two (9.5%) denied that ALTAI would help make their AI more trustworthy.

Summary. Our companies were aware that AI theft is a possibility. Performance checks were mostly performed at least weekly, and performance drops were mostly countered via alarms and retraining. Despite being agnostic to ALTAI, most companies had implemented most of its elements—especially robustness and human oversight.

4.3. Viewpoint on the AI Act

Here, we first focus on two pivotal requirements of the European AI Act: transparency towards the end-users and ethical usage of AI. Then, we reveal the companies' opinions about the AI Act.

[10] One participant reported not having any performance check in place (contradicting their previous assertion) while another one who stated not to have performance checks revised their original statement.

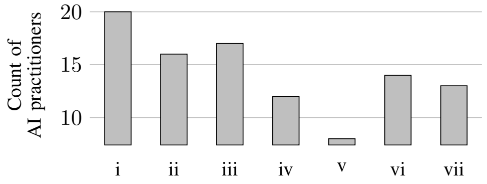

Figure 4. Replies to *Q.6.5*, asking which ALTAI standards are implemented by our companies. *(i)* is technical robustness and safety, *(ii)* human agency and oversight, *(iii)* privacy and data governance, *(iv)* transparency, *(v)* diversity and non-discrimination, *(vi)* environmental and societal well-being, *(vii)* accountability.

Transparency towards end-users. We asked the companies (*Q.5.7*) whether they disclosed to the end-user that they interacted with an AI. We counted 20 replies to this question, as one company stated that their AI does not interact with end-users. Of these, thirteen interviewees (81.3%) did disclose, one (4.8%) planned to do so soon, and two (12.5%) did not disclose that the client was interacting with an AI. We further asked (*Q.6.1*) whether our companies believed that informing their clients about AI increased the trustworthiness of the product. Five interviewees (23.8%) denied this assumption, while sixteen (76.2%) agreed. We further inquired these latter sixteen companies whether it would be feasible to explain their system to an end user—using a scale from 1 ("*not feasible, too complicated*") to 10 ("*feasible, can explain*"). While one interviewee (6.25%) indicated their AI was complicated to explain, seven (43.75%) rated an explanation as rather feasible ($\geq$ 5,$\leq$ 8). Eight (50%) believed that explaining their AI is possible ($\geq$ 9). As these questions implied a certain knowledge from lay people about AI, we inquired whether a company would benefit from more AI education within the general population. To this end, we used a scale from 1 ("*not benefit at all*") to 10 ("*benefit a lot*"). Although the median reply was 7, the replies were rather scattered with seven (33.33%) companies choosing a value $\leq$ 4.

Readiness for the AI Act. We further tried to asses how ready companies were for the AI Act. We first asked (*Q.5.6*) whether they were well prepared to potentially disclose documentation. Seven interviewees (33.3%) estimated that they had the right amount of documentation, whereas eight (38.1%) asserted that they would need to disclose more information than they currently had. Finally, six (28.6%) admitted to not having thought about this issue yet. This reply seemed to be heavily influenced by the companies's sector, with one interviewee saying "*We have enough documentation because of the compliance processes in our [healthcare] sector.*" We further asked (*Q.3.5*) whether companies had already thought about mitigating biases. While two (9.5%) had not, an additional two (9.5%) stated that they planned to address bias mitigation soon. Seventeen (81%) reported already taking bias mitigation into account. We consequently asked (*Q.3.6*) which measures were taken to mitigate bias, allowing to mention more options. Eleven companies used manual bias checks for their data, two applied manual overfitting checks, four relied on a labeling process, four deployed statistical analysis, two used an automated data cleanup process, and three deployed other methods. As the AI Act also requires accountability, we asked in *Q.3.3* whether the companies monitored accountability. Twelve (57.1%) interviewees answered yes, while four (19%) did not, and the remaining five (23.8%) planned to introduce corresponding tests. Lastly, we asked (*Q.3.4*) whether the companies considered explainability or accuracy as more important. Eleven (52.4%) companies said that accuracy was most important, while four (19.0%) preferred explainability. Six (28.6%) chose a trade-off between both properties.

Practitioners' thoughts on the AI Act. We asked (*Q.5.8*) the companies whether they think that regulations and cyber security measures harm their business. Most (sixteen, 76.2%) companies insisted that such regulations added value to their business. Three (14.3%) stated that they had not thought about the topic before and only two (9.5%) argued against the regulations saying that they harmed business. While positive opinions related to the increased trustworthiness, opponents reasoned that "*regulations [...] do not allow us to do certain projects.*" We also inquired (*Q.5.2*) whether they preferred more general or fine-grained regulations. Thirteen (61.9%) were in favor of more fine-grained regulations, four interviewees (19%) had no preference, and another four interviewees (19%) opted for more general. Arguments for more fine-grained regulations included "*give each party [...] more clarity*" or to avoid "*interpret[ing] what the regulations say.*" On the other hand, companies reasoned that "*general regulations are better because technology evolves too fast for fine-grained regulation.*" We also asked an open-ended question (*Q.5.9*) to learn about suggestions how to improve current legislation. Twelve (57.1%) companies wished for more fine-grained regulation, seven (43.75%) suggested that regulation should be designed by AI experts. Two (9.5%) expressed their desire for better communication of regulations, more standard use cases, and a minimal amount of regulation. One (4.8%) participant suggested less strict regulation. Two (9.5%) expressed that they were fine with the regulations as is.

SUMMARY. Most companies were preparing themselves for the AI Act. The companies also mostly implemented transparency with the end-user. However, few prefer explainability over accuracy, while most found documentation to be an issue. Intriguingly, most companies perceived regulation as an opportunity and even wished for more fine-grained regulation, and regulation dictated by AI experts.

4.4. What about the public opinion?

When asked *Q.5.7* (about transparency towards end-users) a participant stated "*having AI is a selling point*", remarking that advertising that their company uses AI is beneficial from a business perspective.

Method. Inspired by such a response, we carried out a survey focused on collecting the opinion of the general population on the usage of the term "AI" by companies. Specifically, after some demographics, we asked three closed questions: (S1) "*Are you knowledgeable of AI?*", with a binary answer; (S2) "*The term* artificial intelligence *is often used by companies to advertise their products in a better light. Do you agree?*", to which possible answers were "Yes/No/I do not know"; (S3) "*Do you think tech companies use the term* AI *in an appropriate way? (According to your own definition of* appropriate*)*", whose answers used a 7-point Likert scale—with 1 being "not appropriate at all" and 7 being "very appropriate". We distributed this survey on social media in Apr. 2023, collecting responses for two weeks.

Findings. We obtained 125 responses. For S1, 88 claimed to be knowledgeable about AI, and 37 believed otherwise. For S2, 85 answered "yes", and 29 answered "no", while 11 answered "I don't know". Intriguingly, for S3, only 10 people responded with a 6 or a 7; 19 answered with a 1 or a 2; while 96 responded with a 3, 4 or 5, with an average of 3.98.

SUMMARY. The respondents of our online survey have mixed views on whether the term "AI" is used appropriately by AI companies.

5. Discussion and Recommendations

Despite some limitations (§3.2), we believe our sample allows deriving intriguing findings that are useful to academics, developers, and lawmakers alike. We thus analyze our findings with respect to existing literature. Then, we state our major takeaways.

5.1. Relation to prior work

We use Table 1 as a scaffold to relate our findings to prior work's insights, before we discuss our results with a focus on e-government. We first focus on **related interview** studies, as systematized in Table 1.

- Meyer and Apruzzese (2022) interviewed smart-grid security practitioners in Europe, and found a skeptical attitude towards *regulation*. In contrast, our sample has a more positive view—likely due to our participants being from smaller companies.
- Some of our findings align with those by Liebl and Klein (2022) who also studied the AI practitioners' opinion on the *AI Act* and found that, e.g., most would not relocate their businesses to other markets (despite the strict European regulation). However, while Liebl and Klein (2022) reported that most AI practitioners believe that creating technical documentation of their AI is relatively easy, our sample thinks otherwise.
- Rakova et al. (2021) reported on the technical difficulty of choosing the right *performance metric*. Our participants similarly struggle with technical details, for example when implementing data deletion.
- Previous interviews in healthcare (V. Johansson et al., 2022) underscored a high need for *transparency and explainability*. Although our sample claimed to implement transparency towards end-users, our interviewees favored performance over explainability—analogous to other industry samples (Bieringer et al., 2022). We further elaborate on these topics below, with a focus on e-governance.
- Rakova et al. (2021) emphasized the need for the right incentives in *bias* avoidance. In contrast, our sample reported a high bias mitigation rate already. A possible explanation is the increased awareness of bias in AI (e.g., due to media reports) which stimulated companies to address this problem directly.
- Our companies were more aware of IP *security* than previously reported (Grosse et al., 2023), even though we did not provide an attack description.

We now discuss the relation of our work to **e-government** with a focus on *explainability*, a crucial property in democracies. According to Madan and Ashok (2022), the tradeoff between accuracy and explainability[11] is crucial in the design of AI systems for public administration. Recently, Leewis and Smit (2023) interviewed 42 Dutch governmental experts on the broad "decision support systems" (not necessarily relying on "neural networks") and found that over 65% believe that explainability is *more important than the "outcome of the decision of a system"*. Furthermore, Wei and Zhou (2023) analyze evidence

[11] Some works use "transparency" and "explainability" (which have different meanings (Endsley, 2023)) almost interchangeably. E.g., Wei and Zhou (2023) use "transparency" to denote the difficulty of AI developers to "*explain* the exact mechanisms behind black box algorithms'. In contrast, we use transparency to denote informing the end-user that there is an AI in place—which aligns with Article 52 (1) of the AI Act. We hence encourage future work to be **more precise**.

of real-world AI failures with a focus on ethics: they underscore that "bad performance" is common in some deployed AI systems, and serious consequences could be prevented with an *increased explainability*. In contrast to both such works, our findings reveal that the companies we interviewed mostly favor sheer accuracy over explainability: a potential reason is that our sample consists mostly of small companies, for which the main goal is to develop "a product that works" and for which explainability is not seen as a priority. On this note, we emphasize that Chouikh et al. (2023) found that most research on AI focuses on the private sector. Nonetheless, as highlighted by Madan and Ashok (2022), the public sector can greatly benefit from AI. However, many "AI tensions" must be overcome, such as a *lack of data literacy* of public administrators—as also reported by our interviewees.

5.2. Implications and future work

We identify practical implications for required resources for companies, their desiderata related to the AI Act, and future avenues for scientific research.

Required resources. Our findings show the need for more infrastructure for the interviewed companies to ensure legitimacy. One example is companies who want to obtain legal counsel are not sure who to turn to (§4.1). We conclude that there is a need for either more personnel or a need for a database that contains personnel with their expertise to ease reaching out for companies in need. Additional required resources concern the information provided about legislation and guidelines. One example here is that the companies did not know about ALTAI (albeit implementing it) and, after learning about the checklist, agreed about its value (§4.2). Analogously, interviewees asked explicitly for better communication of regulation (§4.3). Lastly, we found that companies may require support with performance and bias assessment (§4.2 and §4.3) since many tasks appear to be done manually.

Linking industry & political agenda. We found an overall positive reception of regulations such as the AI Act (§4.3) or guidelines such as ALTAI (§4.2). Yet some companies explicitly expressed their desire for more fine-grained regulations (§4.2). Given some evidence for legal conflicts originating from unclear legislation and regulation (§4.1), companies seemed concerned that the interpretation of the AI Act is ultimately left to court cases instead of upfront, clear definitions and laws. Changing, if at all, the AI Act is however the duty of the European Parliament. We would like to suggest future scientific work instead to improve our knowledge, and thus the basis on which regulations are also made.

Future work. Many of our companies struggled with data deletion procedures, in particular on a large scale (§4.1). This highlights the need for more work in machine *un*learning (Bourtoule et al., 2021) to provide the industry with the necessary, well-understood tools to solve these tasks, in particular when AI is involved. Independent of this, we found that many participants are concerned about their intellectual property in relation to AI (§4.2). Intriguingly, however, many of the participants also stated that they believed their IP was safe (§4.2). More work is needed to assess the severity of AI threats in the wild, and how susceptible real-world AI systems are. Finally, the AI Act (and regulation in general) *is likely to change*: more works similar to this study will be needed in the future to ensure that the AI industry can keep up the pace with legislation.

6. Conclusions

We revealed the viewpoint of 21 European AI companies on the upcoming AI Act. Many companies were preparing for the Act. However, to ease legal compliance, resources such as trained personnel and legal information are needed. Desiderata for the AI Act from industry include more fine-grained regulations and more input from AI experts. Overall, companies' perception of legislation such as the AI Act was positive. In a complementing survey, we found that lay people have mixed feelings towards the way AI companies use the term "AI". Our paper is a call for action: more work is needed in machine unlearning for GDPR compliance, to understand companies' attitudes towards legislation, and to help overall legislation compliance.

Acknowledgements. We want to thank the participants and the anonymous reviewers for the valuable feedback; and the Hilti corporation for funding.

References

Apruzzese, G., et al. (2023). The role of machine learning in cybersecurity. *ACM DTRAP*.

Bieringer, L., Grosse, K., Backes, M., Biggio, B., & Krombholz, K. (2022). Industrial practitioners' mental models of adversarial machine learning. *SOUPS 2022*.

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Patt. Recog.*

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*.

Bohr, A., & Memarzadeh, K. (2020). Artificial Intelligence in Healthcare (Chapter 2).

Bourtoule, L., et al. (2021). Machine unlearning. *IEEE S&P*.

Brundage, M., et al. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation.

Caldwell, M., Andrews, J. T., Tanay, T., & Griffin, L. D. (2020). AI-enabled future crime. *Crime Science*.

Campion, A., Gasco-Hernandez, M., Jankin Mikhaylov, S., & Esteve, M. (2022). Overcoming the challenges of collaboratively adopting artificial intelligence in the public sector. *Social Science Computer Review*.

Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Bus. Horiz.*

Carter, L., Liu, D., & Cantrell, C. (2020). Exploring the intersection of the digital divide and artificial intelligence: A hermeneutic literature review. *THCI.*

Chouikh, A., Khechine, H., & Gagnon, M.-P. (2023). Tertiary study on the use of artificial intelligence for service delivery: A bibliometric analysis of systematic literature reviews. *HICSS-56*.

Dwivedi, Y. K., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *IJIM*.

Endsley, M. R. (2023). Supporting Human-AI Teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*.

European Commission. (2020a). *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*.

European Commission. (2020b). *White paper on artificial intelligence: A european approach to excellence and trust* (White Paper 65 final).

European Commission. (2021). *Artificial intelligence act*.

George, A. S., & George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. *PUIIJ*.

Gragousian, D. (2022). How businesses should respond to the EU's Artificial Intelligence Act. *WEF*.

Grosse, K., Bieringer, L., Besold, T. R., Biggio, B., & Krombholz, K. (2023). Machine learning security in industry: A quantitative survey. *IEEE TIFS*.

Haefner, N., Wincent, J., Parida, V., & Gassmann, O. (2021). Artificial intelligence and innovation management: A review, framework, and research agenda. *Technological Forecasting and Social Change*.

Hagendorff, T. (2020). The ethics of ai ethics: An evaluation of guidelines. *Minds and Machines*.

Hinsen, S., Hofmann, P., Jöhnk, J., & Urbach, N. (2022). How can organizations design purposeful human-AI interactions: A practical perspective from existing use cases and interviews. *HICSS-55*.

Hoffmann-Riem, W. (2020). Artificial intelligence as a challenge for law and regulation. *Regulating artificial intelligence*.

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). Quo vadis artificial intelligence? *Discover Artificial Intelligence*.

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, ai comes—an interview study of organizational ai readiness factors. *Business & Inf. Syst. Eng.*

Jordan, M. I., & Mitchell, T. M. (2015). Machine Learning: Trends, Perspectives, and Prospects. *Science*.

Kaminski, M. E. (2021). The right to explanation, explained. In *Res. handbook on inf. law and governance*.

Königstorfer, F., & Thalmann, S. (2020). Applications of artificial intelligence in commercial banks–a research agenda for behavioral finance. *JBEF*.

Kshetri, N. (2021). Economics of artificial intelligence in cybersecurity. *IEEE IT Professional*.

Laacke, S., Mueller, R., Schomerus, G., & Salloch, S. (2021). Artificial intelligence, social media and depression. a new concept of health-related digital autonomy. *The American J. of Bioethics*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*.

Leewis, S., & Smit, K. (2023). What other factors might impact building trust in government decisions based on decision support systems, except for transparency and explainability? *HICSS-56*.

Liebl, A., & Klein, T. (2022). AI Act impact survey.

Madan, R., & Ashok, M. (2022). Ai adoption and diffusion in public administration: A systematic literature review and future research agenda. *Gov. Inf. Q.*

Meyer, J., & Apruzzese, G. (2022). Cybersecurity in the Smart Grid: Practitioners' Perspective. *ICSS Workshop*.

Mink, J., Kaur, H., Schmüser, J., Fahl, S., & Acar, Y. (2023). "Security is not my field, I'ma stats guy": A Qualitative Root Cause Analysis of Barriers to Adversarial Machine Learning Defenses in Industry". *USENIX Security*.

Mirsky, Y., et al. (2022). The threat of offensive AI to organizations. *Computers & Security*.

Mueller, B. (2021). *How much will the ai act cost europe?* (Tech. rep.).

Mundell, I. (2023). The ecosystem: Sprawling AI Act may deprive European start-ups of Investment.

Ning, H., Yin, R., Ullah, A., & Shi, F. (2021). A survey on hybrid human-artificial intelligence for autonomous driving. *IEEE T-ITS*.

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Educ. Inf. Technol.*

Petersson, L., Larsson, I., Nygren, J. M., Nilsen, P., Neher, M., Reed, J. E., Tyskbo, D., & Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: A qualitative interview study with healthcare leaders in sweden. *BMC Health Serv. Res.*

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A new organizational chassis for artificial intelligence-exploring organizational readiness factors. *ECIS*.

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2021). Where responsible ai meets reality: Practitioner perspectives on enablers for shifting organizational practices. *ACM HCI*.

Reisach, U. (2021). The responsibility of social media in times of societal and political manipulation. *EJOR*.

Rothenberger, L., Fabian, B., & Arunov, E. (2019). Relevance of ethical guidelines for artificial intelligence–a survey and evaluation.

Ruschemeier, H. (2023). AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal. *ERA Forum*.

Stachofsky, J., Schaupp, L. C., & Crossler, R. E. (2023). Measuring the effect of political alignment, platforms, and fake news consumption on voter concern for election processes. *Gov. Inf. Q.*

Stone, P., et al. (2016). Artificial intelligence and life in 2030: The one hundred year study on artificial intelligence. *arXiv*.

Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). Stealing Machine Learning Models via Prediction APIs. *USENIX Security*.

Tricomi, P. P., Facciolo, L., Apruzzese, G., & Conti, M. (2023). Attribute inference attacks in online multiplayer video games: A case study on Dota2. *ACM CODASPY*.

V. Johansson, J., Bentzen, H. B., & Mascalzoni, D. (2022). What ethical approaches are used by scientists when sharing health data? an interview study. *BMC Medical Ethics*.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*.

Wei, M., & Zhou, Z. (2023). AI Ethics Issues in Real World: Evidence from AI Incident Database. *HICSS-56*.

White House Office of Science and Tech. Policy. (2022). *Blueprint for an AI Bill of Rights* (tech. rep.).