The AILA Methodology for Automated and Intelligent Likelihood Assignment

Giampaolo Bella
Dipartimento di Matematica e Informatica
Università di Catania
Catania, Italy
giamp@dmi.unict.it

Cristian Daniele
Department of Digital Security
Radboud University
Nijmegen, The Netherlands
cristian.daniele@ru.nl

Mario Raciti
Osservatorio Astrofisico di Catania
INAF – Istituto Nazionale di Astrofisica
Catania, Italy
mario.raciti@inaf.it

***Abstract*—Risk assessment is core to any institution's evaluation of risk, notably for what concerns people's privacy. The assessment often relies on information stated in a policy shaped as a text document. The risk assessor, or analyst in brief, is called to understand documentation that can be long, unclear or incomplete, hence subjectivity or distraction may strongly influence the process, particularly for identifying each relevant asset and for the assignment of the likelihood value of a given threat to an identified asset. The aim of this paper is to reduce the influence of subjectivity and distraction through risk assessment by means of our methodology for the Automated and Intelligent Likelihood Assignment (AILA). While the analyst's role cannot be emptied, it is facilitated through entities identification and likelihood assignment to threats for assets. The methodology adopts Natural Language Processing for summarisation and entity recognition, it tailors fully-supervised Machine Learning over policy documents and it leverages an existing tool supporting risk assessment, PILAR, in order to gain a more objective likelihood assignment. The paper demonstrates AILA over three real-world case studies from the automotive domain, culminating with the risk assessment exercises over the privacy policies of Toyota, Mercedes and Tesla. The executable components of AILA, the AILA Entity Extractor and the AILA Classifier are released as open source.**

Keywords—policy, risk assessment, natural language processing, likelihood, convolutional neural network, machine learning

I. Introduction

This paper recognises the widespread application of risk assessment to the privacy field, also according to the prescriptions of the General Data Protection Regulation (GDPR). Risk assessment benefits from the expert analyst's policy understanding and evaluation. While it still seems impossible to entirely replace the expert analyst's work, this paper seeks out to automate the analyst's perception of a policy and to reduce the analyst's subjectivity through what perhaps is the hardest step in a risk assessment process: the determination of the likelihood values.

A. Researh Questions

Following ISO/IEC 27005 [1], the risk assessment process rests on the identification of assets and on the definition of potential threats related to each specific asset. For each asset-threat pair, the analyst is called to determine the "chance of something happening" [1], namely a likelihood value, typically on a scale of 1 to 5. The analyst also has to decide the impact of the occurrence of the threat over the given asset. Risk is calculated as a proportional combination of likelihood and impact for the given asset-threat pair. The overarching motivation for our work is that likelihood determination often implies an approximate estimation that may be biased by subjectivity. Such subjectivity may be minimised by using a machine learning model trained on policies labelled by risk assessment experts. One element of subjectivity is the understanding of available policies, documents that may provide useful information, particularly indicating the relevant assets and threats, in support of the entire risk assessment process, also to inform about likelihood and impact of the given threat on the given asset. However, policies are often verbose or incomplete and, in any case, long to read, hence their interpretation may even be subject to the reader's distraction. On such bases, we introduce the following research questions:

RQ1: Can we define a computer-supported process to assist the human analyst through the task of understanding the given policies towards the determination of an entity list and the assignment of a likelihood for each threat that comes from a given list and may affect any of those assets?

The related work shall demonstrate below that this question is currently open. It is looking at a very common scenario where the analyst builds a list of typical privacy threats and is then asked to evaluate them over a target system, say an infrastructure or a software, described through a policy. It would be very useful if a computer-supported process could tell the analyst what entities arise from the policy and also provide a general likelihood indication on whether each asset may be affected by any of the given threats.

RQ2: Can the process mentioned in RQ1 be integrated with or automated through a tool, namely a software application, if this exists, supporting the overall risk assessment process?

The related work shall demonstrate below that also this question is currently open. A few tools exist, and noteworthy is the one that we choose as our main software application due to its maturity and European Commission endorsement, called PILAR [2]. It must be noted, however, that existing tools usually

come with builtin likelihood values associated per threat. This values obviously ignore the features and the niceties of the target system. For this reason, the tools may also challenge the analyst with the task of determining appropriate likelihood values by hand or, if necessary, modifying the predefined ones. PILAR, in particular, allows the analyst to enter modifier values to likelihood values that the tool predefines. Arguably, the modifiers account for target-system specific details, but the challenge for the analyst remains the same.

B. Related Work

The state of the art is conveniently partitioned per topic.

1) Risk Assessment Tools. A few risk assessment frameworks and tools exist and are applicable to various scenarios in different ways. The main supporting tools developed so far are commercial products and services offered by leading companies in the field. Therefore, it is inherently daunting to interrelate all state-of-the-art software. The European Union Agency for Cybersecurity (ENISA) listed an inventory of the most popular RM / RA tools [3], between December 2005 and March 2006. Among these, those that are currently maintained are: PILAR [2], offered by the Spanish Ministry for Public Administrations to support the MAGERIT methodology [4]; and TRICK [5], provided by the private company itrust and compliant with ISO/IEC 27005. Although both are commercial tools, only PILAR can be used without explicit request to the vendor — the user is allowed to download the software as well as to generate a 30-days evaluation license locally. Hence our choice to adopt PILAR as the supporting tools for our methodology and experiments, which remain general and may be applied with other tools.

2) NLP Applications to Privacy Policies. The importance of fully understanding policies and unveiling hidden risks through their analysis has been observed in several papers. The use of Natural Language Processing techniques has helped researchers to develop tools like the Completeness Analyzer by Costante et al. [6], which assigns a degree of completeness to a policy, as the level of completeness is an important aspect to evaluate in the analysis process. Similarly, other studies aim to estimate the extraction of salient features from a policy, such as the automatic categorisation proposed by Ammar et al. [7], or the possibility to quickly identify and understand relevant privacy statements, using text categorisation, as in the contribution by Liu et al. [8]. In addiction, more complex projects and frameworks have been developed, such as POLICIS by Harkous et al. [9], which enables queries on natural language privacy policies, predicting a set of classes for each part of the corpus. This work is useful in understanding the nature of the policy as well as in an automatic annotation of the policy with labels from a prespecified taxonomy. However, none of these works suggest relevant information for the purposes of risk assessment, especially for the determination of the likelihood to a certain asset-threat pair.

3) Fairness in Privacy Policies. Nagpal et al. [10] were among the first to conduct studies related to the fairness of a policy — the fairness level indicates how fair, proper and clean a text is, regarding privacy and security concerns for the users. They proposed a methodology to automatically extract a fairness value from public law documents leveraging semantic relatedness, namely the identification of some form of lexical or functional association between two words or concepts, based on the contextual or semantic similarity of those two words, regardless of their syntactical differences. An inherent limitation is the necessity of manually creating a seed set of WordNet [11] senses, which have to be used as a reference for the similarity. Also, the word vector model, namely a type of word representation that allows words with similar meaning to have a similar representation, may turn out unable to represent the various shades of meaning of the same word. Other challenges arise from those sentences bringing hidden implied meaning, as well from those that are meaningful in a specific domain.

C. Contributions

The overarching contribution of this work is the AILA methodology for the Automated and Intelligent Likelihood Assignment through a risk assessment process based upon information from a given policy and a given list of threats. The first challenge that AILA takes up is the automated recognition of the relevant assets from the policy by means of Natural Language Processing (NLP) techniques. The AILA Entity Extractor is the software component of AILA that addresses this first challenge, and is released open source [12]. The biggest challenge perhaps is the automated assignment of a fairness level to (each statement of) the policy. This remains open due to the inherent risk of errors [13]. A supervised classification could yield the best results but the problem reduces to:

- Having a large dataset for the sake of training a model, as also remarked elsewhere: The most natural way to address this problem could be to learn a supervised classification for sentences. But as discussed earlier, we do not have any publicly available large dataset in the legal domain that has explicitly tagged privacy policies. Thus we are left with the choice of semi-supervised approach [11].
- Having a dataset devoid to any sort of bias.

We managed to obtain a relevant dataset from outputs of the "Terms of Service; Didn't Read" [14] (ToS;DR), which labels a sentence as fair when it will "respect your rights and will not abuse your data" [14]. Therefore, a fairness label signifies the extent to which a policy statement respects natural persons' privacy. The service covers the privacy policies of popular services such as Amazon, Facebook and Wikipedia. AILA leverages that dataset to train a Convolutional Neural Network, hence shifting from semi-supervised to the advocated fully-supervised Machine Learning (ML). Next, AILA adopts the trained model to test relevant privacy policies, namely Toyota, Mercedes and Tesla, and automatically derive the fairness levels of each sentence. AILA's aims are more farsighted, namely to reduce subjectivity through the analyst's task of assigning likelihood values, following the identification of the relevant assets. The AILA Classifier is the software component of AILA that addresses this second challenge, and is released open source [12]. Our target case studies come from the automotive industry and are large car manufacturers' privacy policies. Our choices are Toyota and Mercedes, the first two car manufacturers in Interbrand's 2020 Best Global Brands (BGB) Report [15] (7th and 8th places, respectively, in the overall classification, which accounts for other areas too). We also add Tesla as a third case

study due to its pioneer role on electric cars. We have intentionally chosen policies missing from (ToS;DR) service, to have absolutely guarantee that the eventual particular sentences structure of the training set will not influence our case study evaluation. In a nutsheel, Tesla's average risk likelihood per asset (and, arguably, corresponding risk level) turns out lower than Mercedes's, which in turn is found lower than Toyota's. The methodology also enables the analyst to combine the automatedly assigned likelihood values with those derived from an existing tool from the state of the art, PILAR [2], whose values are purposely ignorant of the target system. To conscientiously combine these two values, it is necessary an in-depth understanding of PILAR's calculation of risk, which we conducted by means of regression analysis, to demonstrate how to practically leverage the AILA likelihood and integrate it with PILAR's calculation of risk levels with the ultimate goal of increasing its realism and reliability. AILA's required inputs are one or more policy documents aimed at regulating assets to achieve an overarching property (and the document is meant to source a risk assessment exercise over the property), a labelled dataset of policies with the same aim and a list of threats to the property. This paper demonstrates AILA for the privacy property over the privacy policies by the three mentioned car brands, using a dataset derived from ToS;DR and considering a list of threats derived from PILAR. The methodology remains general and applicable to different inputs concerning other properties, e.g. cybersecurity or safety.

II. The AILA methodology

AILA addresses the research questions stated above by supporting the analyst during risk assessment towards the automation of:

- the recognition of the relevant asset from a given policy;
- the assignment of likelihood values to the threats related to a given asset, under consideration of (specific details of the target system gathered from) the given policy;
- the assignment of likelihood values to the threats related to a given asset, under consideration of (specific details of the target system gathered from) the given policy via state-of- the-art tools for risk assessment.

A fragment of a simple file management policy is useful as a running example to demonstrate the AILA methodology:

File management policy, North America. To ensure data security, users with different privileges can be created. Any agent can be a Normal User or a Super User. Any agent playing the role of Normal User is Permitted to read the public files. All the public files are stored in the root folder, to be accessible to all the users. Any agent playing the role of Normal User is Permitted to write his own public folders. Each folders can contain only text files. Any agent playing the role of Super User is Obliged to change his password weekly. Any agent playing the role of Normal User is Obliged to change his password monthly. If any agent lost his password, it is possible to request a temporary password which will be valid for 15 minutes from the moment of the request. A Super User can create different Normal Users. Any agent playing the role of Super User is Permitted to read the all files. Any agent playing the role Super User is Permitted to write his own secret file. All the agent names are stored in a special file.

The AILA methodology supports the automated extraction of assets and related fairness values from a policy, then interpreting threat likelihood over an asset in terms of the fairness of the policy over that asset. It consists of three macro steps that are demonstrated.

Step 1. Automated asset extraction). This step removes the parts of the text that are irrelevant to the extraction of the assets, producing a list of entities, from which the analyst — thanks to his experience — may choose the relevant assets. Additionally, each asset is bundled with the sentences that mention the very asset name or a synonym. First, the policy undergoes text summarisation in terms of N- Grams. For the sake of demonstration, the bigrams extracted from our running example are:

[Super, User]

[Normal, User]

[his, password]

[are, stored]

These enter an entity recognition algorithm, which, produces the following outcome over our example:

1) *Super user*
2) *Normal user*
3) *Password*

Then comes the collection of the policy sentences pertaining to the identified entity. With our example, this produces:

1) Super user:
    a. *Any agent playing the role of Super User is Obliged to change his password monthly.*
    b. *Any agent playing the role Super User is Permitted to read the all files.*
    c. *Any agent playing the role Super User is Permitted to write his own secret file.*
    d. *A Super User can create different Normal Users.*
2) *Normal User:*
    a. *Any agent playing the role of Normal User is Permitted to read the public files.*
    b. *Any agent playing the role of Normal User is Permitted to write his own public folders.*
    c. *Any agent playing the role of Normal User is Obliged to change his password monthly.*
    d. *A Super User can create different Normal Users.*
3) *Password:*

a. *Any agent playing the role of Super User is Obliged to change his password weekly.*

b. *Any agent playing the role of Normal User is Obliged to change his password monthly.*

c. *If any agent lost his password, it is possible to request a temporary password which will be valid for 15 minutes from the moment of the request.*

Step 2. Likelihood determination through AILA. This step leverages Machine Learning for the computer-based assignment of a likelihood value. This is feasible through a dataset formed from a corpus of privacy policies where each sentence is labelled by the ToS;DR community to signify the sentence's fairness, namely the extent to which the sentence respects natural persons' privacy. That dataset is leveraged to train a ML model, which can then be applied to a target policy. The findings obtained on our running example are reported as "Fairness per sentence" in Table I. It is then natural to average the fairness values of the sentences related to an asset to derive the fairness value of the asset. This information is conveniently displayed to the analyst, as detailed later, through a navigable HTML page, also supporting manual adjustments that the analyst may want to make to the fairness values. All relevant items are now available to define the likelihood because the given threats are assumed to pertain to the property that the given policy wants to establish on the assets. Targeting the privacy property and leveraging the ML model based upon ToS;DR, one way to define the AILA likelihood on the traditional range 1 to 5 is as the opposite of fairness. The underlying assumption for this choice is that fairness correlates with protection.

Step 3. Combined likelihood determination. The AILA likelihood can be used to continue the risk assessment exercise on a spreadsheet as customary. However, it can be used to better inform the exercise as carried out on PILAR through various methods, for example by replacing the PILAR likelihood by the AILA likelihood or by the average of the two. Moreover, we must understand how Pilar works to evaluate whether and how the risk levels that the tool calculates are influenced by likelihood variations. It turns out that for any non- irrelevant risk level, namely above 2, likelihood linearly influences risk levels, hence we may conclude that likelihood variations are consistently reflected on risk levels.

III. Case studies: Toyota, Mercedes and Tesla

Cars are increasingly complex and interconnected, treating a variety of personal data such as cabin preferences, music preferences, GPS coordinates and sensor data including camera streams. Car manufacturers therefore are data controllers that are called to comply, at least in Europe, with the General Data Protection Regulation. It is thus not surprising that car manufacturers' privacy policies are very developed. We demonstrate the outcomes of AILA on the privacy policies of Toyota and Mercedes, the first two car manufacturers in Interbrand's 2020 Best Global Brands (BGB) Report [15] (7th and 8th places, respectively, in the overall classification, which accounts for other areas too). We also add Tesla's privacy policy as a third case study due to the brand's pioneer role on electric cars. After having obtained the likelihood of the relevant assets thanks to AILA for each of the three car brands, it can be seen that Tesla's average likelihood (and, arguably, corresponding risk level) is medium (or 3, or M), corresponding to an average fairness level of 0,41; Mercedes's average likelihood is high (or 4 or H), corresponding to an average fairness level of 0,26; Toyota's average likelihood is very high (or 5 or VH), corresponding to an average fairness level of 0,14.

TABLE I. Outcome of AILA Step 2 and AILA Step 3 on our running example

Asset	Sentence	Fairness per sentence	Fairness per asset	AILA Likelihood (AILA Step 2)	PILAR Likelihood	Combined Likelihood (AILA Step 3)
(a)	(i) (ii) (iii) (iv)	1.0 0.20 1.0 0.50	0.67	2	3	2
(b)	(i) (ii) (iii) (iv)	0.0 1.0 1.0 0.50	0.62	2	1	1
(c)	(i) (ii) (iii)	1.0 0.80 0.70	0.83	1	3	2

The total 57 assets for the three car brands along with their likelihood values are not presented here due to space limitations but are available online [12].

IV. Conclusions

This paper advanced AILA, an innovative methodology to reduce human subjectivity through risk assessment, and applied it to the assessment of given threats related to privacy. However, AILA is general for any risk assessment exercise relying on likelihood assignment upon the basis of information stated in prose in a policy document. AILA responds somewhat positively to the stated research questions, which pertain to how to automate the entity — asset, after the expert validation — extraction process from a policy, how to automate the likelihood

assignment to given threats for those assets and how integrate the above with a tool of the state of the art. AILA's main software components, the AILA Entity Extractor and the AILA Classifier are released open source to promote the widespread development of the area. AILA's integration with PILAR is conceptually simple now that we understand how the latter calculates impact and risk levels, but cannot be completed because PILAR is not open source. The application of AILA to the automotive field was profitable on all three case studies, Toyota, Mercedes and Tesla, showing how to reduce a few thousand words to only a few dozen entities, hence facilitating asset extraction dramatically. AILA also conveniently automated the assignment of the likelihood values for all assets, offering significant reduction of subjectivity. A limitation of AILA is that it can be used only when a non biased and labelled dataset is available in order to train the ML model; still, this is an inherent limitation of ML in general. Future work includes deeper semantic analysis of the policies to tune the likelihood values, which are currently assigned per asset, more finely per asset and per threat. Another useful direction would be to write an open-source risk assessment tool from scratch to truly integrate PILAR with AILA. Open sourcedness would spark off a community of developers as well as additional research, and the new AILA tool could become the (at least de facto) standard tool for risk assessment.

ACKNOWLEDGMENT

We thank Dandelion support team, for providing us a non-profit plan, and José A. Mañas, for the precious suggestions about PILAR and in general risk assessment. This research is supported by the project MEGABIT - Piano di incentivi per la Ricerca di Ateneo 2020/2022 (PIACERI) – linea di intervento 2, DMI - University of Catania.

REFERENCES

[1] Organización Internacional de Normalización. ISO/IEC27005: Information technology-Security techniques -Information security risk management. ISO, 2008.

[2] "Ear/pilar. "https://pilar-tools.com/en/index.html.

[3] "Enisa rm/ra tools." https://www.enisa.europa.eu/topics/threat-risk-management/risk- management/current- risk/risk- management-inventory/rm- ra-tools.

[4] "Magerit. "https://www.ar-tools.com/magerit/index.html.

[5] "Trick." https://www.trickservice.com.

[6] E.Costante, Y.Sun, M.Petković, andJ.denHartog, "A machine learning solution to assess privacy policy completeness: (short paper)," in Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society, WPES '12, (New York, NY, USA), p. 91–96, Association for Computing Machinery, 2012.

[7] W.Ammar, S.Wilson, N.Sadeh, and N.A.Smith, "Automatic categorization of privacy policies: A pilot study," School of Computer Science, Language Technology Institute, Technical Report CMU-LTI-12-019, 2012.

[8] F.Liu, S.Wilson, P.Story, S.Zimmeck, and N.Sadeh, "Towards automatic classification of privacy policy text," School of Computer Science Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-ISR-17-118R and CMULTI-17-010, 2018.

[9] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin, and K. Aberer, "Polisis: Automated analysis and presentation of privacy policies using deep learning," in 27th {USENIX} Security Symposium ({USENIX} Security 18), pp. 531–548, 2018.

[10] R.Nagpal, C.Wadhwa, M.Gupta, S.Shaikh, S.Mehta, and V.Goyal, "Extracting fairness policies from legal documents," 2018.

[11] "Wordnet. "https://wordnet.princeton.edu.

[12] B. for review, "Aila source code," 2021. https://anonymous.4open.science/r/AILA-source-code-32AB/README.md

[13] M.ROEHLING, ""extracting" policy from judicial opinions: The dangers of policy capturing in a field setting," Personnel Psychology, vol. 46, no. 3, pp. 477–502, 1993.

[14] "Term of services; didn't read." https://tosdr.org.

[15] "Interbrand's 2020 best global brands (bgb) report. "https://www.interbrand.com/best- global- brands/.

[16] Samany you Gard, "Tldr this." https://tldrthis.com.

[17] Autosummariser.com, "Automatic text summarizer," 2013. https://autosummarizer.com.

[18] Tools 4 noobs, "Online summarize tool," 2007. https://www.tools4noobs.com/summarize/.

[19] N.M.Alsharmanand, I.V.Pivkina, "Generating summaries through unigramand bigram: Text summarization," International Journal of Information Technology and Web Engineering (IJITWE), vol. 15, no. 1, pp. 64–74, 2020.

[20] E.Villatoro-Tello, L.Villaseñor-Pineda, and M.MontesyGómez, "Usingword sequences for text summarization," in International Conference on Text, Speech and Dialogue, pp. 293–300, Springer, 2006.

[21] N.AndhaleandL.Bewoor, "A nover view of text summarization techniques," in 2016 International Conference on Computing Communication Control and automation (ICCUBEA), pp. 1–7, IEEE, 2016.

[22] "Dandelion." https://dandelion.eu/.

[23] "Dandelion entity extraction api. "https://dandelion.eu/docs/api/datatxt/nex/v1/.

[24] J. Wei and K. Zou, "Eda: Easy data augmentation techniques for boosting performance on text classification tasks," arXiv preprint arXiv:1901.11196, 2019.

[25] S.Albawi, T.A.Mohammed, andS.AlZawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), pp. 1–6, Ieee, 2017.

[26] "Magerit book i - the method." https://www.pilar-tools.com/doc/magerit/MAGERIT_v3_book_1_method_PDF_NIPO_630-14-162-0.pdf.

[27] "Pilar glossary of terms. "https://www.pilar-tools.com/en/glossary/index.html.