Institution: University of Essex

Unit of Assessment: 11 – Computer Science and Informatics

Title of case study: Development of a platform for civilian-led reporting of human rights violations in conflict zones improving human rights outcomes across the Middle East and North Africa

Period when the underpinning research was undertaken: 2012-2020

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Udo Kruschwitz,	Professor	1997-2019
Massimo Poesio	Professor	2000-2017
Ayman Alhelbawy	Research Officer	2014-2020
Chris Fox	Reader	2002-2020
Jon Chamberlain	Lecturer	2017-2020

Period when the claimed impact occurred: 2016 –2020

Is this case study continued from a case study submitted in 2014? N

1. Summary of the impact

Essex NLIP research resulted in the development of an online platform for secure real-time civilian-led reporting of human rights violations in conflict zones providing more accurate monitoring in areas inaccessible to traditional human rights monitoring missions. International human rights organisation, Minority Rights Group International (MRG) and the Ceasefire Centre for Civilian Rights, deployed the platform between 2017 and 2019 in Iraq, Iran, Syria, the Middle East and North Africa. In Iraq more than 3,700 cases of human rights violations were documented using the platform since 2017. The platform has enabled MRG to grow its activities and was essential for the establishment and operation of Ceasefire, furthering the practice of civilian-based monitoring of violations. This secure and easy to use platform has empowered and enabled ordinary civilians to document violations, mitigating the risks associated with reporting of human rights abuses in conflict situations. The evidence of human rights violations obtained from the platform has enabled documentation of claims for reparations, to compensate the victims of human rights abuses (cases documented on the platform can form the basis for reparations claims under the Iraqi national Law 20 reparations system, which has paid out over USD 355 million in compensation) and for international oversight, generating reports which have been used by the United Nations (UN), ultimately improving human rights outcomes for civilians.

2. Underpinning research

Essex Natural Language Information Processing (NLIP) researchers collaborated with MRG and the Ceasefire Centre for Civilian Rights (Ceasefire hereafter) from 2014 to design and implement a new platform to support civilian-led monitoring of human rights violations [G1]. Their challenge was to develop a reporting platform which would enable civilian-led reporting in inaccessible areas which would be secure, easy to use and be able to assess the accuracy of reports, supporting both English and Arabic languages. Developing a platform for this purpose was particularly challenging as Arabic is an under-resourced language in NLP and requires a unique set of skills. The research of the NLIP team, on information extraction from social media, Arabic natural language processing and crowdsourcing addressed these challenges.

The Essex research provided insights into crowdsourcing which focused on determining a "gold standard" based on a large number of individual statements with varying degree of reliability e.g., [R1]. The reporting platform was a development of Ushadhidi, an open-source crowdsourcing platform used to map reports of violence, but its security model was insufficient for Ceasefire's needs. The Essex researchers developed a new security model and user access control, providing additional structural support for collaborating organisations with additional structural and security modifications [R2]. Essex's Arabic natural language processing toolkit (AraNLP) [R3] together with methodology to identify named entities (e.g., names of people and places) which describes how names, places and dates can be extracted [R4], provides the basis of the Arabic text processing pipeline that automatically processes social media texts to identify potential human rights

violations. AI-powered approaches for classifying massive amounts of data into pre-defined categories need enough reliable training data and machine learning algorithms of sufficiently high classification quality. This was addressed by Essex research on extracting and managing content from social media, using crowdsourcing techniques to label and evaluate the collected data and the framework for using crowdsourcing to create language resources for under-resourced languages such as Arabic [R5] [R6]. The Essex language resource of Arabic texts representing different types of violence including human-rights violations [R7] was the first such resource. This addressed the lack of previous work on detecting human rights violation in English and Arabic, the modest accuracy of existing methods and the paucity in robust and scalable resources to process Arabic natural language. One reason for this is the difficulty in applying methods developed for languages such as English to Arabic, a highly derivational and inflectional language that is characterised by a complex set of morphological features including gender, number, person, case, state, mood, and voice. In addition, Arabic has a set of clitics, which attach to the stem after affixes such as conjunctions, prepositions, future marks, definite articles, and pronouns. [R2] reports Essex's insights into mining and classifying Arabic Twitter in order to identify potential human rights abuse incidents in a continuous stream of social media data within a specified geographical region. Essex's analysis showed deep learning approaches such as Long Short-Term Memory (LSTM) classification precision was 84% with an F1-score of 75%. The practical usefulness of the Twitter mining tool for the analysts' work was also demonstrated and was deemed of high enough quality to be used in the practical setting [R2].

This Essex research resulted in a new platform for Ceasefire (Fig.1), described in [R2], which monitors human rights violations within a specified geographic area via (i) a structured reporting system which experts and individual witnesses submit to, enabling civilian-led monitoring, combined with (ii) AI mining of social media yielding a continuous stream of data which have been classified as signals of potential violations within the same region. Essex's system constituted the portal first portal of its kind which combines (i) and (ii). The platform's sources include reports submitted by users directly to the website, email and SMS, Twitter, and mainstream news reports, in both Arabic and English [R2]. Reports submitted through the form are stripped of personally identifying information and plotted onto a live, interactive map visible on the tool's landing page, which shows the distribution of violations by location and type. Sensitive personal data is stored securely on a secondary server in case it is needed for future follow-up or court proceedings. The reports obtained are made available to partner organisations and the public at the same portal web interface where reports are submitted.
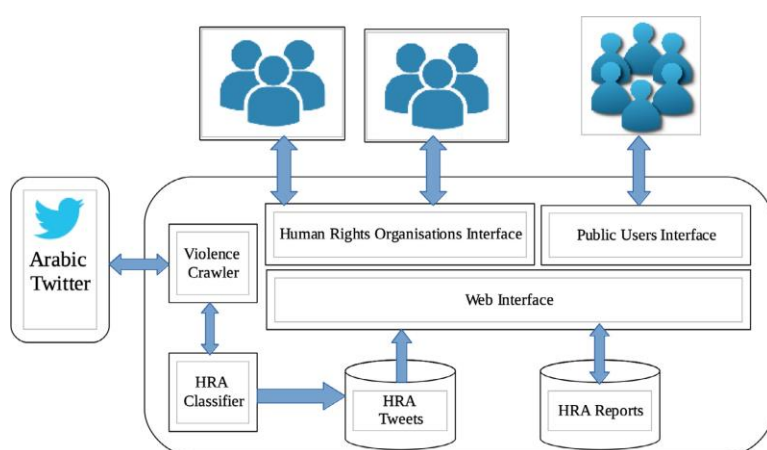

Fig. 1: Ceasefire: a framework for reporting and monitoring of human rights .

3. References to the research [can be supplied by HEI on request]

This research was published in respectable books, journals and conferences with global reach. All papers were peer-reviewed and are internationally recognised for the originality, significance and rigour of the research.

[**R1**] T. Fornaciari and M. Poesio. 2014.Identifying fake Amazon reviews as learning from crowds.

In Proceedings of EACL, Gothenburg, April 26-30 2014. https://dx.doi.org/10.3115/v1/E14-1030

[**R2**] A. Alhelbawy, M. Lattimer, U. Kruschwitz, C. Fox, and M. Poesio. 2020. An NLP-Powered Human Rights Monitoring Platform. Expert Systems with Applications, Volume 153, 113365. https://doi.org/10.1016/j.eswa.2020.113365.

[**R3**] M. Althobaiti, U. Kruschwitz and M. Poesio. 2014. AraNLP: a Java-based Library for the Processing of Arabic Text. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC), Reykjavik, Iceland, 2014. http://www.lrec-conf.org/proceedings/lrec2014/pdf/621_Paper.pdf

[**R4**] M. Althobaiti, U. Kruschwitz, and M. Poesio. 2015. Combining Minimally-supervised Methods for Arabic Named Entity Recognition. Transactions of the ACL (TACL), 3:243–255, 2015. https://doi.org/10.1162/tacl_a_00136

[**R5**] M. Poesio, J. Chamberlain and U. Kruschwitz. Crowdsourcing. In N. Ide and J. Pustejovsky (eds.), Handbook of Linguistic Annotation (Springer), 2015. ISBN 978-94-024-0881-2

[**R6**] M. El-Haj, U. Kruschwitz, and C. Fox. 2015. Creating language resources for under-resourced languages: methodologies, and experiments with Arabic. Language Resources and Evaluation. 49 (3), 549-580. https://doi.org/10.1007/s10579-014-9274-3

[**R7**] A. Alhelbawy, M. Poesio, and U. Kruschwitz. 2016. Towards a Corpus of Violence Acts in Arabic Social Media. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC), Portoroz, Slovenia, 2016. http://www.lrec-conf.org/proceedings/lrec2016/pdf/553_Paper.pdf

4. Details of the impact

Developing and implementing a secure reporting platform enabling civilians to document human rights violations

The ability of traditional human rights monitoring via investigative missions to work effectively and achieve their aims is challenged by restricted access to insecure territories, as well as cost, time, questions over politicisation, oversight, legitimacy and representation. With advances in communications technology, civilian-led monitoring in conflict zones has become a complementary, and sometimes the principal, source of information on human rights violations [S1]. Although new technologies provided hardware for civilians to report violations in theory, obtaining reliable human rights data and information that can be documented required a new system. Essex NLIP researchers collaborated with MRG (a leading international human rights organisation, with over 150 partners across 60 countries) [S2] and Ceasefire (founded in 2014 as an in-house centre of specialist expertise on civilian-led monitoring within MRG) on the 2014-2017 Ceasefire Programme to develop and implement such a system of civilian-led monitoring of human rights abuses [S3].

The Executive Director (ED) of Ceasefire (and former ED of MRG until 2018) confirms that '***Essex's NLIP research was applied to develop the essential technological infrastructure to support civilian-led monitoring***, *including dual Arabic/English language reporting, as well as techniques for social media data mining. Specifically, they* ***developed an integrated online platform*** *to receive and process information from sources, including reports submitted by users directly to the website, email and SMS, information posted on Twitter, and mainstream news reports, in both Arabic and English which was critical to the Ceasefire programme's success*' [R2] [S3]. They add: '***Essex's specialist insights in Arabic NLP, working in areas such as text summarisation and named entity extraction, provided the essential foundations for developing the platform***' [e.g. R3] and '*through further research in a project funded by MRG and*

Innovate UK, the NLIP researchers constructed the online Ceasefire platform [...] Essex's NLIP research was the essential basis for that project's results' [R1-7] [S3]. The Ceasefire platform comprised two technological developments: A violations reporting form, allowing users to submit reports of violations directly to the website and the addition of social media data mining tools to the website, allowing information on human rights violations to be automatically harvested from Arabic language Twitter feeds [R2] [S3]. Importantly '***Essex's NLIP research provided the capability to adapt and enhance the systems for new environments***, *and in line with changing modes of communication'* so that the reporting platform is '*applicable to different contexts, and conflict situations'* [S3]. The development of **this new platform was** '***essential for the success of the Ceasefire programme which has empowered civilians subject to violations in highly insecure environments'*** [S3].

This reporting platform was piloted by MRG and Ceasefire in Iraq in 2016 with MRG's partner organisations, including Asuda - Combating Violence against Women, then publicly rolled out in Iraq in 2017, Iran in 2018, part of Syria in 2019 and more widely across the Middle East and North Africa in 2019 [S3]. Accessible to anyone with an internet connection, **the platform empowers and enables ordinary civilians to document violations of their rights in a way that conforms to international legal standards and increases the future usability of the data** [S3, S4 p.57]. The creation of a safe, secure and easy to use platform [R2] has helped to **mitigate the risks to those reporting human rights abuses in conflict situations**, including civilians, journalists and humanitarian agencies [S4]. '*In the current situation in Iraq, the major risk facing human rights activists and defenders is reporting violations [...] The online tool helps them prevent these risks as they easily use the tool and it is anonymous.'* Director of an Iraqi NGO [S4 p 58]. For civilians, the platform enables them to report incidents at any time and in a more confidential way than talking with a representative of an NGO, which could be under surveillance [S4]. In Iraq more than **3,700** cases of human rights violations were documented using the platform since 2017 [S3, S4 p. 58], several hundred in Syria since 2019 – '***a considerable body of documentary evidence of violations'*** [S3]. In addition to the violations reported via the platform, **information harvested from social media alerts the relevant authorities to violations, providing a means of early warning** [S3]. '*The Twitter crawler also provides real-time information on the popular reaction to reports of violations, giving us a perspective on people's attitudes to human rights abuses, including at local level'* [S3].

Contributing to the work of UN bodies, increasing the availability of reparations and improving human rights outcomes
Data obtained from the reporting platform contributes evidence for influential reports [e.g. S5a, b] submitted by MRG/Ceasefire to UN bodies for international oversight of human rights issues. Ceasefire's ED states '***Different UN mechanisms use the information provided as a result of the tool*** *and some have referred to our reports which the platform provided essential evidence for'* [S3]. For example, *'our work on Iran, based partly on one of the tools, was quoted by the UN Secretary-General in his report on Iran for the UN General Assembly in 2019* [see S6 p15]. *This is the most significant and authoritative annual report on human rights in the country'* [S3]. When protests erupted in Iraq in 2018, reports of excessive use of force by security forces against peaceful protesters began to appear on platform. These cases were incorporated into a wider report on repression of civilian activists in Iraq [S5a] which was then used as a basis for urgent appeals and advocacy pushes at the UN Human Rights Council and other forums [S4]. Ceasefire have a strong record of having their recommendations on human rights adopted: *'we made a total of 63 separate recommendations on Iraq to the UN Universal Periodic Review, the UN Committee on the Elimination of Racial Discrimination and the UN Committee on the Elimination of Discrimination against Women; of these, 28 were fully adopted and 16 partially adopted by the relevant UN body'* Ceasefire ED [S3].

The evidence of human rights violations obtained from the platform has also **enabled documentation of claims for reparations, to compensate the victims of human rights abuses** [S3]. '***Cases documented on the platform can form the basis for reparations claims under the Iraqi national Law 20 reparations system***, *which has paid out over USD 355 million in compensation. This aided our cooperation with the UN assistance mission in Iraq and with the Iraqi*

Parliament: we subsequently formed an agreement with the Iraqi Human Rights Committee to provide technical assistance on the drafting of new human rights legislation. We also have a developing relationship with IOM (International Organization for Migration) because they have taken a particular interest in reparations.' Ceasefire ED [S3]. By enabling the development of the platform, the Essex NLIP research has ultimately '***empowered civilians subject to violations in highly insecure environments and boosted the availability of reparations. This had a major effect on human rights outcomes***' Ceasefire ED [S3].

Supporting MRG's human rights operations and the establishment of Ceasefire furthering the practice of civil society based monitoring of violations
Through the development of the new platform, Essex research enabled MRG '***to develop a major new area of expertise***' [S7] and *'provided the technology for the system to enable civilian-led reporting of human rights abuses which Ceasefire's work depends upon and was thereby **an essential factor in establishing and maintaining Ceasefire, furthering the practice of civil society based monitoring of violations**. This in turn helped Ceasefire and MRG to secure significant new external funds for related projects, including the award of a new EUR1,000,000 project by the EU'* [S3], [S7, p4]. MRG is financed by donations and grants from foreign governments, trusts and foundations, and from individual donations [S7, p2]. The increase in its income from a year before the project began to 2017 was almost entirely attributed to the platform (S7, p5-6, items 9 and 10, pages 5-6). MRG's staff acquired new skills in the management of online reporting platforms, improved understanding of user needs and behaviour in the reporting of human rights violations on the ground, better appreciation of the adaptation of international standards on human rights monitoring and reporting, as well as a wider awareness of digital advances to support the human rights movement generally [S7 p5]. Furthermore, '*training in the use of the reporting platform has been undertaken on two occasions for civil society organisations in Iraq*' [S7 p5]. MRG also employed a new Programme coordinator to implement the results of the project [S7 p8].

5. Sources to corroborate the impact

[S1] Eyes on the Ground: Realizing the potential of civilian-led monitoring in armed conflict (2017)
[S2] Minority Rights Group Website https://minorityrights.org/about-us/
[S3] Testimonial from Executive Director (and former Executive Director of MRG), Ceasefire
[S4] 50 Years of Minority Rights Group Report (2020)
[S5] a. 'Civilian Activists under Threat in Iraq' report (2018) b. 'Cultivating Chaos: Afrin after Operation Olive Branch' report (2020)
[S6] UN General Assembly, Report on Human Rights in the Islamic Republic of Iran, 8 February 2019, A/HRC/40/24
[S7] Final Report on Knowledge Transfer Partnership KTP009488