ESTABLISHED 1987

UNITED KINGDOM REPORT

PRIVACY LAWS & BUSINESS

DATA PROTECTION & PRIVACY INFORMATION WORLDWIDE

AI-powered Onfido one of first selected for the ICO's Sandbox

Onfido, an identity verification company, will research how to identify and mitigate algorithmic bias in machine learning models used for remote biometric identification. By **Ali Vaziri** of Lewis Silkin LLP.

In the digital economy, identity is the key to unlocking access to services widely relied on in order to participate in society. Since in-person interaction is no longer always required of, or expected by, users, the challenge faced by many online organisations is how to know a person wanting to access their services is who they claim to be, and in a

Continued on p.3

Smart-home study weighs the privacy risks involved

Martin Kraemer and **William Seymour** at the University of Oxford report on an ICO-funded research project investigating how 'smart' doesn't have to mean invasive.

Studies and media reports about smart home technologies and smartphone apps show that consumers have little awareness of the information they expose to companies, advertisers, and other cohabitants when they use these services. These thought processes of how devices (and the information economy more generally) work can leave users feeling

Continued on p.6

Issue 105 SEPTEMBER 2019

UNITED KINGDOM
report

ISSUE NO 105 **SEPTEMBER 2019**

PUBLISHER
Stewart H Dresner
stewart.dresner@privacylaws.com

EDITOR
Laura Linkomies
laura.linkomies@privacylaws.com

DEPUTY EDITOR
Tom Cooper
tom.cooper@privacylaws.com

REPORT SUBSCRIPTIONS
K'an Thomas
kan@privacylaws.com

CONTRIBUTORS
Ali Vaziri
Lewis Silkin LLP

Martin Kraemer and William Seymour
University of Oxford

Charlotte Reddish, Jenai Nissim and Alison Deighton
Hello DPO

Emma Hughes and Nicola Fulford
Hogan Lovells

Lore Leitner and Christopher Foo
Wilson Sonsini Goodrich & Rosati

PUBLISHED BY
Privacy Laws & Business, 2nd Floor,
Monument House, 215 Marsh Road, Pinner,
Middlesex HA5 5NE, United Kingdom
Tel: +44 (0)20 8868 9200
Email: info@privacylaws.com
Website: www.privacylaws.com

Subscriptions: The *Privacy Laws & Business* United Kingdom Report is produced six times a year and is available on an annual subscription basis only. Subscription details are at the back of this report.

Whilst every care is taken to provide accurate information, the publishers cannot accept liability for errors or omissions or for any advice given.
Design by ProCreative +44 (0)845 3003753
Printed by Rapidity Communications Ltd +44 (0)20 7689 8686

ISSN 2047-1479

comment

Data protection issues on and around Brexit

It has not been discussed much in the general media what a detrimental impact a no-deal Brexit would have on data transfers and international business, even if it was recognised as one of the top issues in the negotiations between the EU and the UK. However, in the Operation Yellowhammer papers, the government also highlights the worst possible scenario for data flows; it warns that an adequacy assessment could take years, and law enforcement data and information sharing between the EU and UK will be disrupted.

A leaked government document suggests that the prime minister has instructed government departments to share data they collect about usage of the GOV.UK portal, without informing individuals. This data would feed into Brexit preparations.

A government spokesperson has told Buzzfeed, which broke the story (www.buzzfeed.com/alexspence/boris-johnson-dominic-cummings-voter-data), that "individual government departments currently collect anonymised user data when people use GOV.UK. The Government Digital Service is working on a project to bring this anonymous data together to make sure people can access all the services they need as easily as possible. No personal data is collected at any point during the process, and all activity is fully compliant with our legal and ethical obligations."

In this issue we report on work that Friends of the Earth has done to make sure that its privacy policy is understandable to everyone (p.8) and why Onfido has embarked on the ICO's Sandbox programme (p.1). Another ICO initiative is its grants programme – read on p.1 about privacy issues with smart homes.

The ICO's new cookies policy has raised some questions (p.16) – not least among international business as there are some differences between that and guidance from France's regulator, the CNIL.
Our correspondents also look at issues about consent, contractual necessity and legitimate interests when using AI (p.20) and how to assess data protection risk (p.12).

Laura Linkomies, Editor
PRIVACY LAWS & BUSINESS

Contribute to PL&B reports

Do you wish to contribute to *PL&B UK Report*? Please contact Laura Linkomies, Editor (tel: +44 (0)20 8868 9200 or email: laura.linkomies@privacylaws.com) to discuss your idea, or offer to be interviewed about your organisation's data protection/Freedom of Information work.

Onfido ... from p.1

way that protects that person's privacy without unduly impacting on user experience.

Onfido enables its clients to verify the identities of their users by processing information readily available to most individuals: images of their identity document and a live face. "Typically, we find that clients want to use Onfido's identity verification (IDV) services to introduce a greater element of trust in their own platform or services – for example to verify the identity of a data subject before letting them rent a property – or in relation to a specific legal obligation imposed on the client. For example, financial institutions need to verify the identity of a data subject as part of their know your customer/anti-money laundering supply chain," Neal Cohen, Onfido's Director of Privacy, told Ali Vaziri in an interview for *PL&B*.

Personal data are generally collected by Onfido either from clients via an API integration or directly from data subjects where a client has integrated Onfido's image capture user flow into their mobile application or website. Onfido then uses machine learning models, supplemented by human review experts, where needed, to determine the likelihood that:

- the identity document is genuine and not fraudulent; and
- the facial image on the identity document matches the image of the data subject's live face and is not a spoofed or fraudulent image.

If the identity document is believed to be genuine and the two faces match, then the data subject has proven that they are who they claim to be.

Onfido's clients receive a report with Onfido's findings in these two steps which they use to make a determination as to whether to onboard the data subject or not. "Clients can configure the service to their own risk tolerance and, where a data subject is not successful, clients can use the report to understand what went wrong and provide recourse to the data subject. Sometimes, this is as simple as asking a data subject to submit a new image because the original image had glare or another defect. Other times, the data subject may have evidenced actual signs of fraud – in which case, the client will likely not admit that data subject to their platform," Cohen says.

By using the data obtained when providing its IDV services to clients to train its machine learning models, Cohen points out that Onfido is able to achieve high levels of performance in the real world as its training data reflects its real world inputs. "Achieving high levels of performance requires large volumes of data that both samples the intrinsic underlying distribution of an individual's facial characteristics and provides realistic representations of extrinsic real world variations. In Onfido's facial recognition models, the intrinsic factors that determine the underlying distribution are the facial attributes of the data subjects (e.g. facial hair, skin type, shape and texture of the facial features, skin colour and tone). The extrinsic factors are the real world variations in the data such as how the photos were taken (e.g. sensors, distance, lighting, and angle)."

The problem

"We have detected that our machine learning models do not always react to all data subjects with the same level of performance. In particular, we have observed this phenomenon in our facial recognition models which perform differently for individuals with different skin tones and other identifying facial features. We believe that this may be due to the low data diversity in our training data, amongst other variables," Cohen explains.

The problem of bias in machine-learning is far from being unique to Onfido and there are many similar reports of other AI technologies performing differently with different types of individuals. As Elizabeth Denham also observed in a recent blog post:[1] "… facial recognition systems are yet to fully resolve their potential for inherent technological bias; a bias which can see more false positive matches from certain ethnic groups."

"In the ICO Sandbox, we would like to pay special attention to this particularly troubling issue," Cohen says. "Our aim is to investigate and improve our machine learning models to operate with less algorithmic bias, particularly in relation to data subject diversity - including where such diversity may be attributable to ethnicity, with a view to potentially extending this to age, gender, and other elements of diversity in the future."

When asked about the specific data protection issues which prompted Onfido's Sandbox application, Cohen explains: "We are trying to understand how a new technology fits under a new data protection framework, and we have a series of fundamental questions. When is personal data (such as facial images and biometrics) a special category of personal data when training machine learning models, given that such processing is not intended to uniquely identify an individual or produce any impact on that individual? Rather, the processing is intended to address data diversity challenges and improve the technology for the benefit of the next person who is to use that technology. Nonetheless, there is the reality that such processing will require data labelling as to the diversity of the data. Following that question, we want to better understand which lawful basis applies. Considering that some technologies such as facial recognition technology require millions, if not billions, of data samples to make the technology effective and accurate, it is difficult to see how the existence of such technology is compatible with an opt-in consent. If that is true, what other safeguards and controls can be introduced to balance the impact on the rights and freedoms of the data subject? We think these are not only legal questions, but ethical questions as well."

Like other AI service providers, Onfido is, where appropriate, using personal data not just in order to provide the requested service but also to develop the underlying technology, including by training Onfido's machine learning algorithms and its human experts. "Questions of data classification and lawful basis are then all the more complicated when considering that even though Onfido is a data processor when providing identity verification services to its clients, Onfido is likely a controller when developing the underlying identity verification technology for the benefit of Onfido and all clients. As a controller with no direct relationship with the data subject, how can we ensure transparency and control for the data subject, and how do we structure our contracts with our clients to reflect the dual nature of being a processor and a controller? There is not much precedent or industry norm, and we have found that clients expect to see the data processing obligations stipulated by Article 28 of the GDPR, which are reflective of a pure processor relationship. Clients also generally have little appetite for Onfido interacting with their data subjects."

The plan

Onfido intends to explore different machine learning research techniques to measure and mitigate algorithmic bias in its machine learning models. In parallel, Onfido also intends to explore the explainability of its machine learning models as well as their robustness and adaptability. Onfido anticipates four stages to its Sandbox plan: (1) assigning labels to datasets to reflect the data diversity attributes it will seek to test and improve; (2) testing its existing machine learning models to identify performance differences towards different types of data (e.g., bias in the data); (3) experimenting with the different research techniques; and (4) testing and monitoring the performance of the refined machine learning models. The nuances of the test plan are currently being discussed with the ICO.

Opportunity to engage with the ICO

Had Onfido not been accepted in the Sandbox, Cohen is adamant that the business would have continued, regardless, in its efforts to find a solution to the issues raised given that, in its view, there is too much at stake to be complacent. "When I heard about the Sandbox, I thought it would be an ideal way to engage with a set of people at the ICO who I believe would be committed to understanding the technology behind our products. That understanding is crucial if a pragmatic and ethical solution to these issues is to be found."

From Cohen's perspective, the value in the Sandbox is to bring together tech/research and policy/law/ethics into a common forum to discuss and resolve issues. He considers that those groups are far too often not in the same room and are having very different conversations about very similar issues. "My aim – and I think this is also the aim of the ICO – is to bridge these worlds," Cohen says.

In addition to the obvious product benefits from its IDV services treating all individuals fairly and equitably – something Onfido is deeply committed to – the business is equally keen to aid the ICO in its understanding of complex technological issues and in the production of regulatory guidance by making its technology, knowhow, and research team available to the ICO. "Since the GDPR has come into force, a number of organisations have pulled back on how they use personal data and put in place overly restrictive compliance structures which, whilst inherently well intentioned, may in practice risk going too far and could stifle innovation and growth," Cohen observes.

"We hope that any learnings will also be shared with other regulators as the issues we are looking to address in the Sandbox transcend disciplines and borders," Cohen says. The UK Centre for Data Ethics and Innovation (CDEI) – established by the Department for Digital, Culture, Media & Sport to provide the Government with advice on the ethical and innovative deployment of data and AI – is to observe Onfido's experience in the ICO Sandbox from the perspective of its own work programme, which includes investigating the issue of algorithmic basis in various sectors with a focus on bias against characteristics protected under the Equality Act 2010.

Cohen, who is also a Technology and Human Rights Fellow at the Carr Center for Human Rights Policy at Harvard University, is leveraging his experience at Onfido to research the legal and ethical challenges of building and using AI for identity verification. Subject to the ICO's consent, Cohen intends to include his experiences in the ICO Sandbox in his larger report on IDV that will be published by the Carr Center. In publishing his findings and experiences in the Sandbox, Cohen aims to empower others that are also seeking to address the legal and ethical challenges of creating AI.

"The Sandbox is also an opportunity to engage with a regulator with considerable resources and international influence – regardless of Britain exiting the EU – and to help shape policy with a knock-on effect internationally. The true value in the Sandbox is the ability to solve critical issues and export the thinking behind those solutions to other jurisdictions in an effort to create a harmonized regulatory environment in which we, and other AI companies, can operate."

The risks in taking part

Sandbox entry is, however, not without risk for Onfido. "The combination of new technology and new laws means that there are a lot of unknowns. However confident we are in our position, in the absence of regulatory guidance, ultimately nothing is certain and there is an inherent risk in sticking your neck out – as Onfido is doing here." Whilst the prospect of a letter of negative assurance or comfort from enforcement was a draw, Cohen is clear that those adaptive mechanisms were far from being the main drivers for participation – not least because, although Onfido is headquartered in the UK, as a global business with global clients they would be of relatively limited value in any event.

"The business knows that there is a possibility that the ICO could switch us off or require a fundamental change to how we do business. Despite this, the most frequent pushback internally was based on a misunderstanding of the word "sandbox". In the tech world, a "sandbox" is a testing environment, and several people thought we would have to transfer data to an environment physically controlled by the ICO – which set off a lot of red flags in terms of security. This is, of course, absolutely not the case. None of the data we process will be disclosed to the ICO or any other party due to us participating in the Sandbox. Disclosure of confidential or commercially sensitive information to the ICO – a body subject to FOIA – was a consideration, but the business is relatively optimistic. We took advice, particularly on how participation might affect our patents, and we do not anticipate disclosing any of our code in any event.

Onfido's decision to apply to enter the ICO Sandbox was carefully considered and many months of work preceded the submission of its formal application. "We have been invested in this process for almost a year now –

well before the Sandbox applications formally opened. There have been a number of informal touch points, the first of which was our response to the ICO's call for evidence last autumn. We also subsequently attended the Sandbox workshop earlier this year, which provided a useful opportunity to ask questions and to solicit feedback from the ICO. The ICO was always receptive to our requests for more information and was very willing to join us on several calls. Through our interactions with the ICO, we were able not only to identify which product we thought would likely be most appropriate for the Sandbox but also to keep refining our proposal once that product had been identified."

The process was resource-intensive and, for an organisation such as Onfido which is growing rapidly and has relatively limited spare resource, Cohen describes it as a gamble: "If we were not accepted into the Sandbox, then a lot of time and effort may have been squandered. Though, of course, the work in preparing for the Sandbox did help escalate our internal thinking on the issues, which is of great value, regardless as to whether we were admitted into the Sandbox. At no point along the way did we receive an indication from the ICO about how well we were doing. That said, we did, however, take some encouragement from the fact that they continued to engage with us."

Although Onfido is now one of the ten successful applicants (from an initial list of 64), the challenges are far from over. Cohen explains that there is a long journey ahead: "While the Sandbox is a great opportunity to bridge two worlds that often do not see eye to eye, this implicitly means that there is a gap that needs to be overcome. Tech companies and privacy regulators do not necessarily operate or think in the same ways, but we need to learn how to work together better and grow from those experiences. Only through mutual cooperation and understanding can we begin to achieve real progress. I am confident that the ICO also shares in those views, and this is what makes the Sandbox such a unique and special opportunity."

AUTHOR

Ali Vaziri is a Managing Associate at Lewis Silkin LLP specialising in information law and media litigation. Email: ali.vaziri@lewissilkin.com

REFERENCE

[1] ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/07/blog-live-facial-recognition-technology-data-protection-law-applies/ (9 July 2019)

ICO SANDBOX PROJECTS

The ICO has selected 10 projects from the 64 applications received and expects the projects to finish by September 2020. The participants are:

FutureFlow: FutureFlow is a RegTech start-up designing a Forensic Analytics platform that monitors the flow of funds in the financial system. Its platform enables multiple financial institutions, regulators and agencies to leverage each other's intelligence on Electronic Financial Crime without heavy reliance on Personally Identifying Information. This collaborative approach to tackling financial crime opens the prospect of higher detection rates with lower false positives, while reducing the burden of scrutiny on each individual and business consumer.

Greater London Authority: In order to reduce levels of violence in London, the Mayor has set up a Violence Reduction Unit (VRU) which is taking a public health approach to this issue. As part of this work, the VRU needs to better understand how public health and social services can be managed to prevent and reduce crime, with a focus on early intervention. There is increasing interest from the VRU, the Mayor's Office of Policing and Crime (MOPAC) and the Greater London Authority (GLA), for health, social and crime data to be looked at in an integrated and collaborative way.

Heathrow Airport Ltd: Heathrow Airport's Automation of the Passenger Journey programme aims to streamline the passenger journey by using biometrics. Facial recognition technology would be used at check-in, self-service bag drops and boarding gates to create a seamless experience for passengers travelling through the airport. Current processes require passengers to present different forms of documentation, such as boarding cards and passports, at different points in their journey to prove their identity and show that they are authorised to travel. By offering passengers the option of using facial recognition technology. they would have the choice to enjoy a frictionless journey through the airport.

Jisc: Jisc is developing a Code of Practice with universities and colleges wishing to investigate the use of student activity data to improve their provision of student support services. This will help them protect both privacy and wellbeing. Jisc provides UK universities and colleges with shared digital infrastructure and services, such as the superfast Janet Network.

MHCLG: The Ministry of Housing, Communities and Local Government's project partners with Blackpool Council and the Department of Work and Pensions, and seeks to match personal information controlled by multiple parties in order to create a dataset that will allow MHCLG to understand more about the private rented sector in Blackpool, who lives there, and how we can help improve the quality of properties.

NHS Digital: NHS Digital is working on the design and development of a central mechanism for collecting and managing patient consents for the sharing of their healthcare data for secondary use purposes, including medical research and regulated clinical trials.

Novartis Pharmaceuticals UK Limited: Novartis is exploring the use of voice technology within healthcare. Through its Voice Enabled Solutions project, Novartis is working with healthcare professionals to design solutions to make patient care easier, and addressing the data privacy challenges posed by this emerging technology.

Onfido: See p.1

Tonic Analytics: The Galileo Programme was launched in 2017 and is jointly sponsored by the National Police Chiefs' Council and Highways England. Galileo's primary focus is on the ethical use of innovative data analytics technology to improve road safety while also preventing and detecting crime.

TrustElevate: TrustElevate provides secure authentication and authorisation for under-16s. TrustElevate is the first company globally to provide verified parental consent and age checking of a child. It is working to enable companies to comply with regulatory requirements, and to make the Internet a safer environment for children, facilitating a more robust digital ecosystem and economy.

Join the Privacy Laws & Business community

The *PL&B United Kingdom Report*, published six times a year, covers the Data Protection Act 2018, the Freedom of Information Act 2000, Environmental Information Regulations 2004 and Privacy and Electronic Communications Regulations 2003.

PL&B's United Kingdom Report will help you to:

Stay informed of data protection legislative developments.

Learn from others' experience through case studies and analysis.

Incorporate compliance solutions into your business strategy.

Learn about future government/ICO plans.

Understand laws, regulations, court and tribunal decisions and what they will mean to you.

Be alert to privacy and data protection law issues and tech developments that will affect your compliance and your reputation.

Included in your subscription:

1. Six issues published annually

2. Online search by keyword
Search for the most relevant content from all *PL&B* publications and events. You can then click straight through from the search results into the PDF documents.

3. Electronic Version
We will email you the PDF edition which you can also access via the *PL&B* website.

4. Paper version also available
Postal charges apply outside the UK.

5. News Updates
Additional email updates keep you regularly informed of the latest developments in Data Protection, Freedom of Information and related laws.

6. Back Issues
Access all *PL&B UK Report* back issues.

7. Events Documentation
Access UK events documentation such as *PL&B Annual International Conferences*, in July, Cambridge.

8. Helpline Enquiry Service
Contact the *PL&B* team with questions such as the current status of legislation, and sources for specific texts. This service does not offer legal advice or provide consultancy.

privacylaws.com/reports

> "Given the rate of change in law, regulation and business practice, it is essential to have concise and up to date information. *PL&B* is always relevant and continues to offer great value."
>
> **Adam Green, Chief Risk Officer, Equiniti**

International Report

Privacy Laws & Business also publishes *PL&B International Report*, the world's longest running international privacy laws publication, now in its 33rd year. Comprehensive global news, currently on 165+ countries, legal analysis, management guidance and corporate case studies on privacy and data protection, written by expert contributors

Read in more than 50 countries by regulators, managers, lawyers, and academics.

Subscriptions

Subscription licences are available:

- Single use
- Multiple use
- Enterprise basis
- Introductory, two and three years discounted options

Full subscription information is at privacylaws.com/subscribe

Satisfaction Guarantee

If you are dissatisfied with the *Report* in any way, the unexpired portion of your subscription will be repaid.