

Citizen Scientists
Investigating Cookies
& App GDPR Compliance

Complying with the general data protection regulation (GDPR):

How hard is it to design privacy-by-default?

Dr. Huma Shah

Assistant Professor

School of Computing, Electronics & Mathematics



"The Internet wasn't built to track people across websites: ... everything you do online is logged in obscene detail"

(Manjoo, 2020)

Today's talk

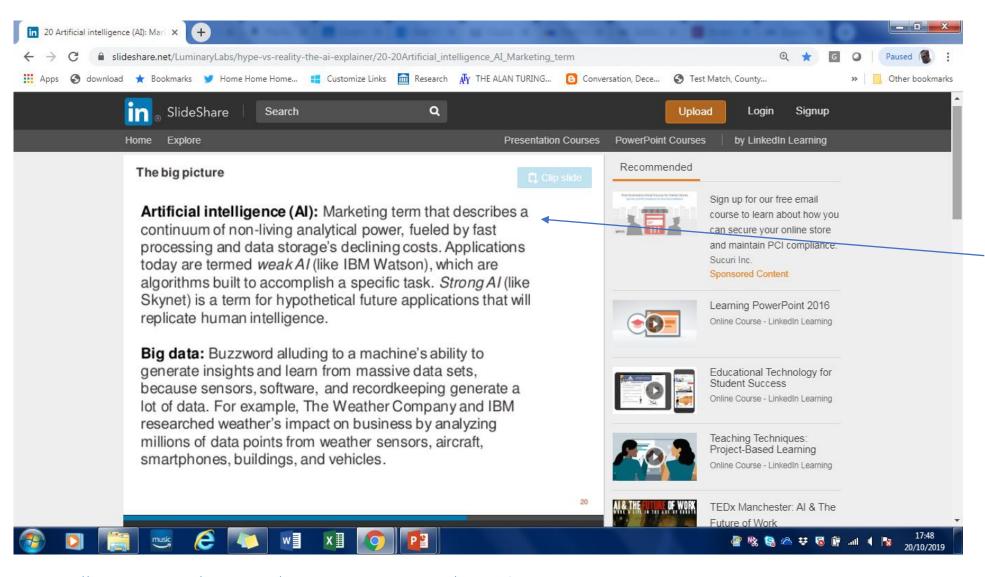
- Quick introduction about my previous work
- A brief look at the extent of the online privacy problem
- Relation to current project: EU H2020 CSI-COP R&I project
- €1.999m 30month; Jan 2020-June 2022

My Activities include....

- Science Director (CoI), EU Horizon2020 SwafS15-2019 project *CSI-COP* (Coventry University coordinating €2m eleven-partner international consortium in grant agreement phase)
- EU Horizon2020 project proposal Reviewer
- International project proposal Reviewer
- Reviewer Q1 journal Computers in Human Behavior
- Member of the EU AI Alliance
- Member of the European Platform for Women Scientists (EPWS)
- West London STEM Ambassador

Book: published by Cambridge University Press, 2016





Al is not a marketing term!!!
It is a scientific field



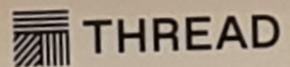
Cheat. Dress better.

Get new ideas for what to wear this autumn.

We combine expert stylists with clever A.I. to
find great clothes in your style, size and budget.

It's so easy it feels like cheating.

Shop over 500 brands and get £15 off your first order at thread.com/tube



Clever AI?

Ad on Metropolitan Line tube, Thursday October 17, 2019

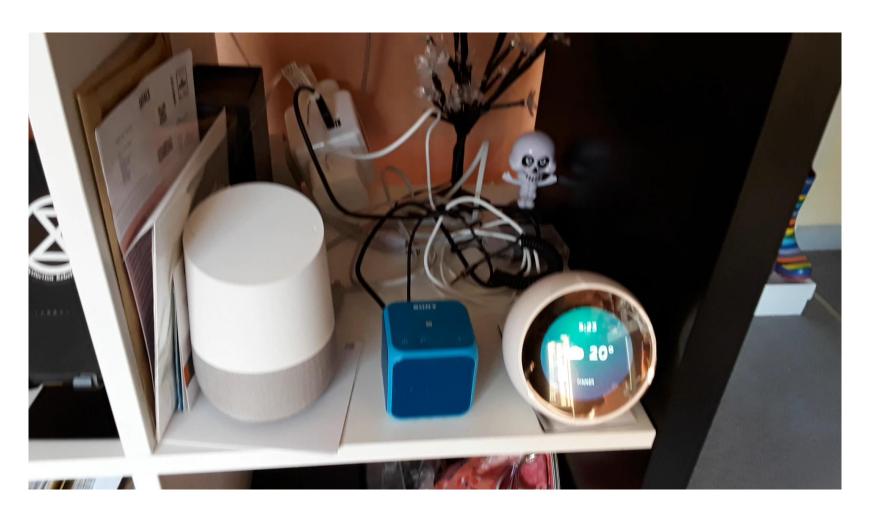
Artificial Intelligence meets Natural Stupidity

Article by **Drew McDermott**

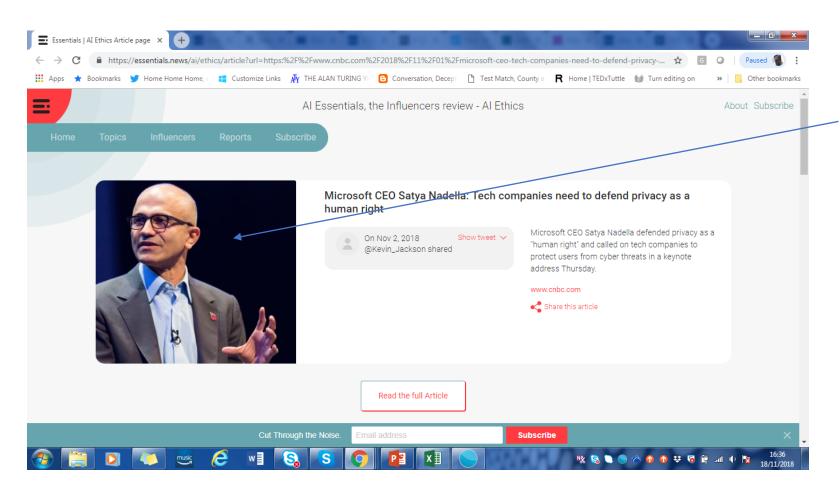
published in

ACM SIGART Bulletin Issue 57, April 1976: pages 4-9

Relevance of Turing's Question-answer test



Privacy is a human right



Microsoft CEO Satya Nadella:

- "Tech companies need to defend privacy as a human right"

Microsoft Privacy Statement

"Microsoft collects data from you, through our interactions with you and through our products. You provide some of this data directly, and we get some of it by collecting data about your interactions, use, and experiences with our products. The data we collect depends on the context of your interactions with Microsoft and the choices you make, including your privacy settings and the products and features you use.
 We also obtain data about you from third parties."

https://privacy.microsoft.com/en-US/privacystatement#mainnoticetoendusersmodule

Privacy as a fundamental right

- "Privacy is a <u>fundamental right</u>:
 - Essential to autonomy
 - Protection of human dignity
 - Serving as the foundation upon which many other human rights are built."

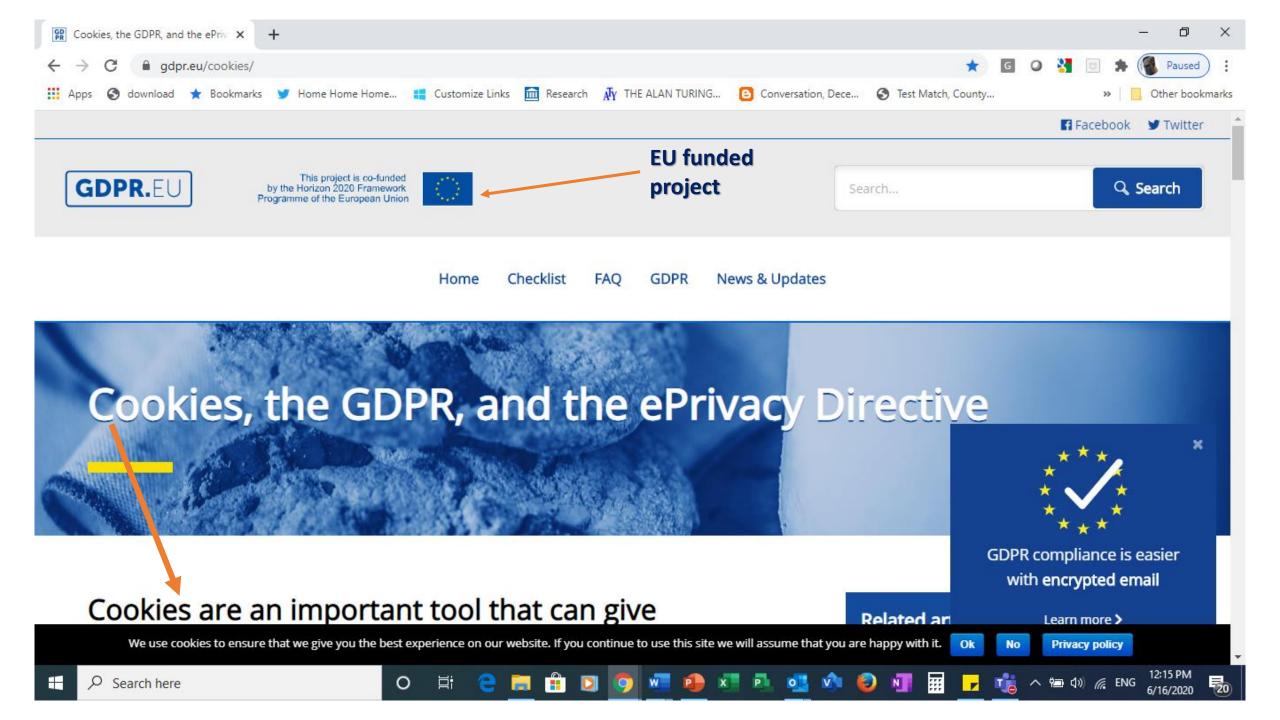
https://privacyinternational.org/explainer/56/what-privacy

Logging by Cookies

Digital cookies:

- "... are small text files that websites place on your device as you are browsing";
- "... can store a wealth of data, enough to potentially identify you without your consent";
- " ... are a primary tool that <u>advertisers use to track your online</u> <u>activity</u> so that they can target you with highly specific ads"

Link to GDPR EU Horizon2020 funded project for further information: https://gdpr.eu/cookies/





The website uses a minimal build of Google Analytics, a service which transmits website traffic data to Google servers in the United States and allows us to notice trends to improve the user experience on our website. This minimal build processes personal data such as: the unique User ID set by Google Analytics, the date and time, the title of the page being viewed, the URL of the page that was viewed prior to the current page, the screen resolution, the time in local timezone, the files that were clicked on and downloaded, the links clicked on to an outside domain, the type of device, and the country, region, and city.

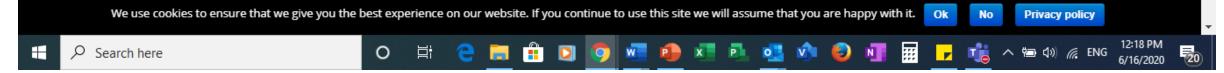
You may opt out of this tracking at any time by activating the "Do Not Track" setting in your browser.

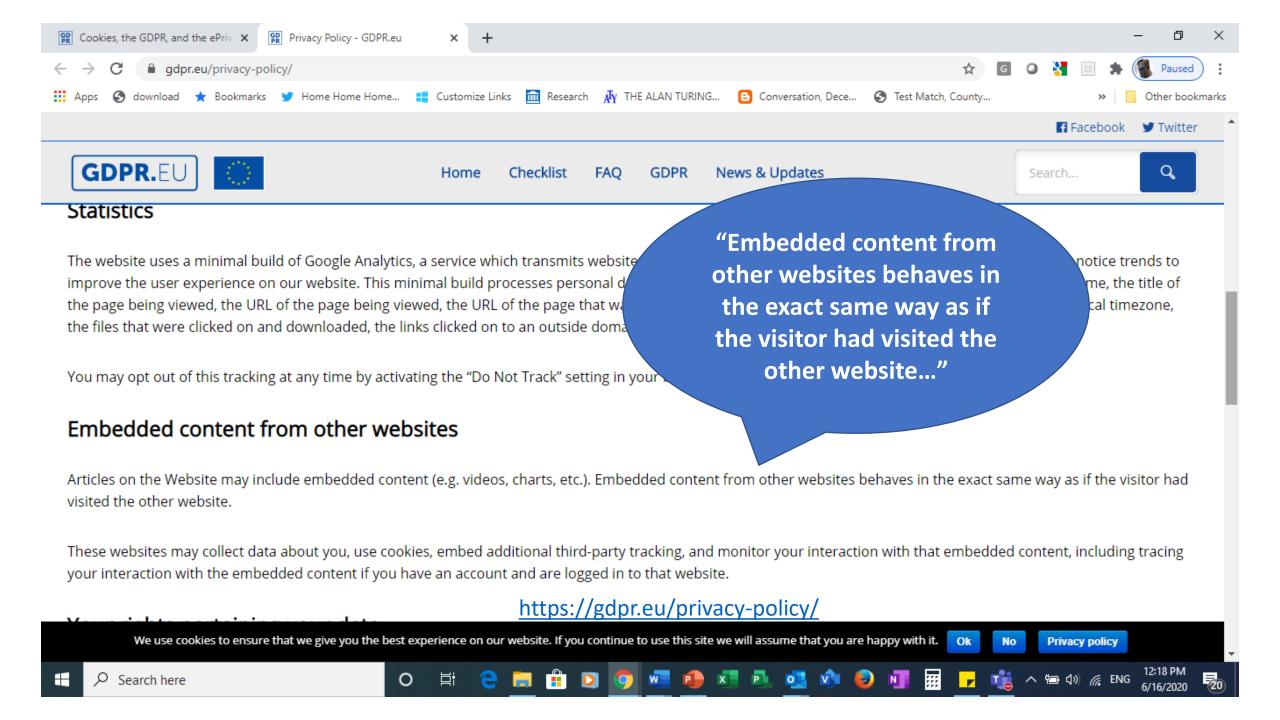
Embedded content from other websites

Articles on the Website may include embedded content (e.g. videos, charts, etc.). Embedded content from other websites behaves in the exact same way as if the visitor had visited the other website.

These websites may collect data about you, use cookies, embed additional third-party tracking, and monitor your interaction with that embedded content, including tracing your interaction with the embedded content if you have an account and are logged in to that website.

https://gdpr.eu/privacy-policy/





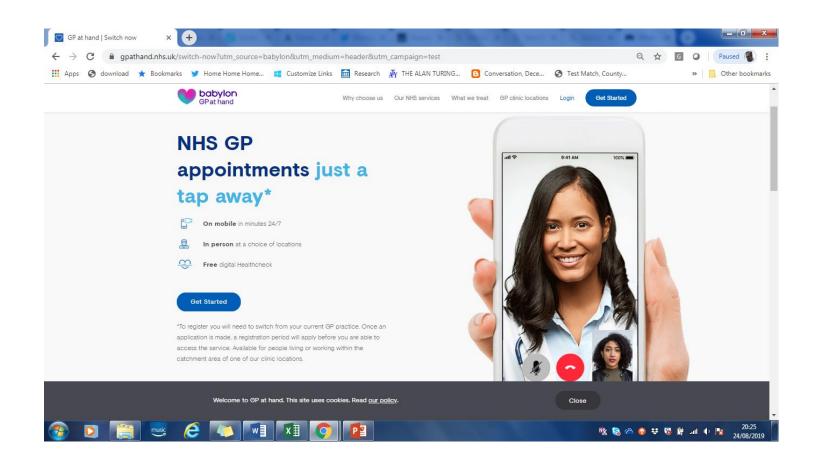
GDPR & Informed Consent

- Case study I give to my students:
 - Royal Free Hampstead NHS Trust collaboration with Google's DeepMind
- Powles and Hodson (2017) investigated first healthcare deals of Google's British-based artificial intelligence subsidiary, DeepMind Technologies Limited, in the period between July 2015 and October 2016.
 - "Google, the world's largest advertising company, that has long coveted the health market"

Data Protection Officers and Data Protection

- Powles and Hodson (2017)
 - "... amount of data transferred [millions of identifiable personal sensitive medical records] is far in excess of the requirements of those publicly stated needs.."
 - "None of the millions [the NHS Trust's patients] in the dataset were either informed of the impending transfer to DeepMind, nor asked for their consent"
 - "We do not know—and have no power to find out—what Google and DeepMind are really doing with NHS patient data, nor the extent of Royal Free's meaningful control over what Google and DeepMind are doing"

Privacy and Smart Phone Apps



Babylon's "mission is to put an accessible and affordable health service in the hands of every person on earth".

Babylon health app:

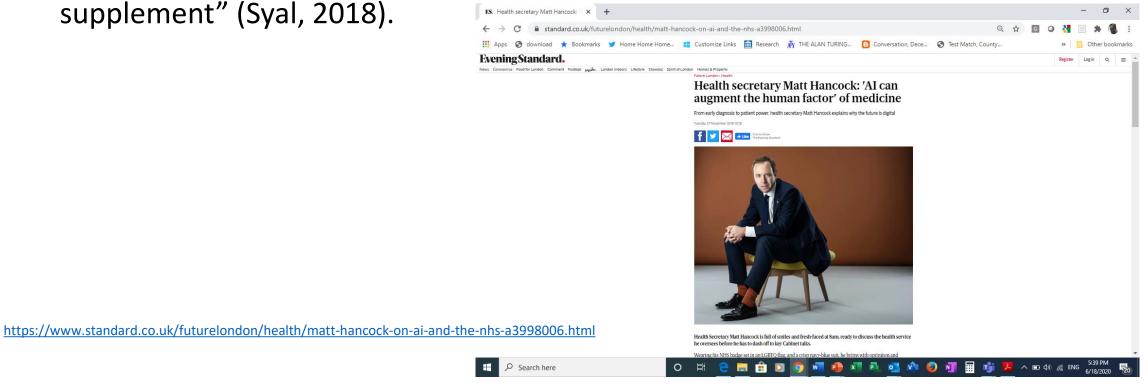
https://www.babylonhealth.com/

Babylon app and the UK Health Secretary

Matt Hancock, current UK Secretary of State for Health:

• Endorsed Babylon "private healthcare company in a sponsored newspaper

supplement" (Syal, 2018).



Babylon app data breach

- "... one of its [Babylon app] users discovered he had been given access to dozens of video recordings of other patients' consultations." (Leo Kelion, BBC: 2020)
- "A follow-up check by Babylon revealed a small number of further UK users could also see others' sessions." (Leo Kelion, BBC: 2020)

Effect of Babylon app data breach

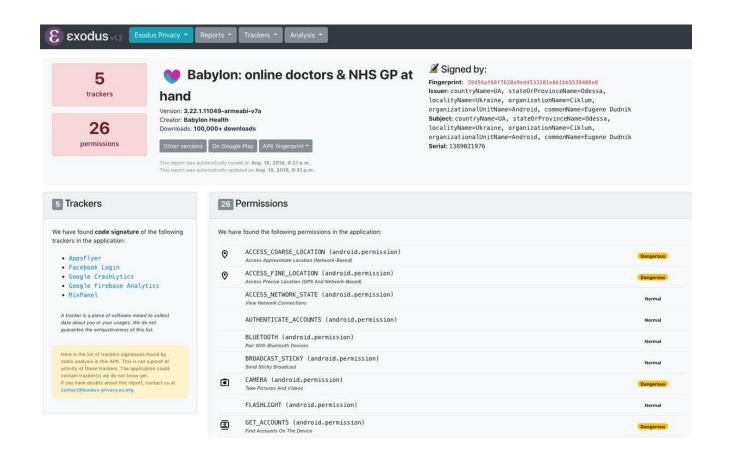
- "Babylon allows its members to speak to a doctor, therapist or other health specialist via a smartphone video call and, when appropriate, sends an electronic prescription to a nearby pharmacy. It has more than 2.3 million registered users in the UK."
- "Leeds-based Rory Glover had access to the service via his membership of a private health insurance plan with Bupa, one of Babylon's partners."

Effect of Babylon app data breach contd.

- "... when he [Rory Glover] went to check a prescription, he **noticed** he had about 50 videos in the Consultation Replays section of the app that did not belong to him."
- "Clicking on one revealed that the file contained footage of another person's appointment."
- "I was shocked," he told the BBC. "You don't expect to see anything like that when you're using a trusted app. It's shocking to see such a monumental error has been made."

Leo Kelion, 2020: https://www.bbc.co.uk/news/technology-52986629

Embedded Trackers in Smart Phone Apps



Babylon health app has these embedded trackers:

Facebook Login

AppsFyer
Mixpanel
Google Firebase
Google Crashlytics

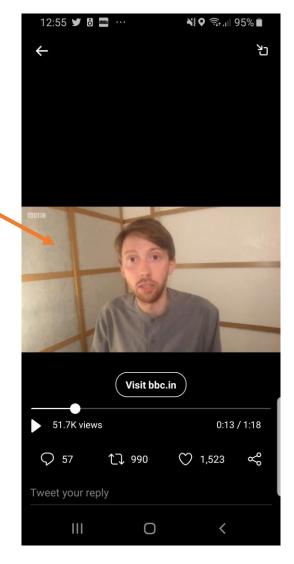
UK Daily Briefing Thursday 18 June

- Matt Hancock makes a comment about Apple re the UK government contact tracing app:
 - '...a "technical barrier" was hit when testing the NHS app on the Isle of Wight... Android phones (Google platform) were far more compatible with it than Apple products.'
 - https://www.lbc.co.uk/news/watch-live-the-governments-coronavirus-press-conference/
 - Isle of Wight trial used a **centralised app:** picked up only 4% of Apple iPhone users (Channel 4 news, 18 June 2020- Andy Davies' item) !Apple's architecture philosophy is privacy-preserving (**Michael Veale** UCL & Turing Institute).

Privacy & contact tracing apps

- Veale's team developed the technology that Google and Apple adapted into the decentralised approach to their joint contact tracing app.
- Veale's findings "already made it into university syllabi" (@mikarv 19/06.20)

https://www.ucl.ac.uk/laws/people/dr-michael-veale



UCL's Dr Michael Veale on BBC Newsnight: 18 June 2020

Sensitive Personal Information and Privacy

- Menstruation data is being monetised:
 - "Out of 36 apps tested by *Privacy International 61% automatically transfer data to Facebook the moment a user opens the app."
 - "This happens whether the user has a Facebook account or not, and whether they are logged into Facebook or not"
 - "Menstruation apps are not just concerned with menstruation cycles"
 - <u>"Menstruapps How to Turn Your Period Into Money (For Others)</u>, they collect information about your health, your sexual life, your mood and more .."

^{*} https://www.privacyinternational.org/long-read/3196/no-bodys-business-mine-how-menstruation-apps-are-sharing-your-data

Internet for tracking?

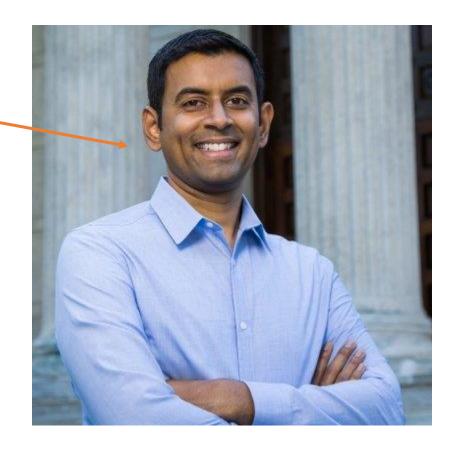
Dr. Arvind Narayanan leads

Princeton's Web Transparency & Accountability Project:

"The web is a cesspit of surveillance"

https://www.cs.princeton.edu/~arvindn/

https://webtap.princeton.edu/



Relation to CSI-COP project?

- EU H2020 (SwafS) programme funded research and innovation idea on an essentially AI ethics project
- 30-month 11-partner project: CU, UPAT, NaTE, TIL, UOULU, BIL, TDL, CTU, Stelar, IB, UAB: https://cordis.europa.eu/project/id/873169
- Leverage citizen science methodology to coinvestigate tracking on websites and in smart phone apps to create taxonomy of trackers

Citizen Scientists in CSI-COP

- Citizen scientists are volunteer enthusiasts who collaborate on real-world projects supporting research of professional scientists
- Citizen science contributes new knowledge through a variety of projects across many disciplines: space science, earth science, etc., see SciStarter:

https://scistarter.org/

Informal Education in Online Privacy

- Free-to access and attend informal education in online privacy and informed consent:
 - MOOC
 - Workshops
- Outcome:
 - Cohort of CSI-COP citizen scientists engaged as a grass-roots movement motivated to investigate cookies in websites and on apps

Research -> Innovation

- Trained CSI-COP citizen scientists will record cookies they find in websites they visit, and apps they use
- Co-develop taxonomy of different types of cookies and embedded trackers (session, functional, necessary, performance, advertising, tracking, etc.)
- **Co-innovate** online knowledge resource an accessible repository of trackers (think *Panama Papers*, or *UCL's Legacies of British slave-ownership online database*)

CSI-COP work completed

• Research:

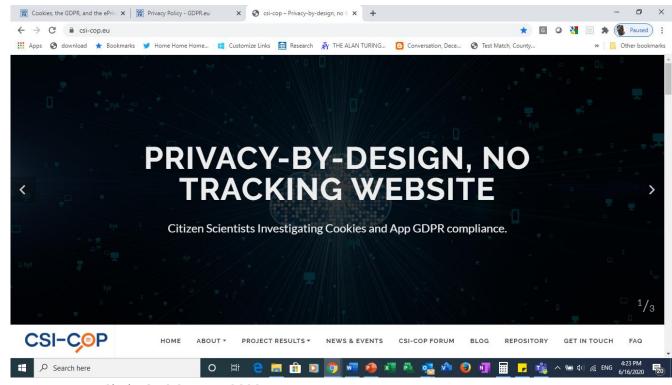
- <u>Completed</u>: produced/submitted open-access report (D2.1 M04 30 April 2020) on **best practices** in engaging, sustaining motivation, and **citizens' science** legacy. DOI: 10.5281/zenodo.3899478
- <u>Completed</u>: submitted initial data management plan (DMP) (D1.7 M06 18 June 2020)

CSI-COP current research tasks

- <u>Completed</u>: exhaustive research to learn how previous citizen science projects have addressed challenge of inclusivity: balanced participation from gender, digital divide, socio-economic, geographical, immigrants/refugees (D2.2 due M06 30 June 2020)
- Underway: creating framework for inclusive CSI-COP citizen science engagement – due M09 (September 2020)

CSI-COP innovation work so far

- Innovation:
 - Completed: project website: https://www.csi-cop.eu



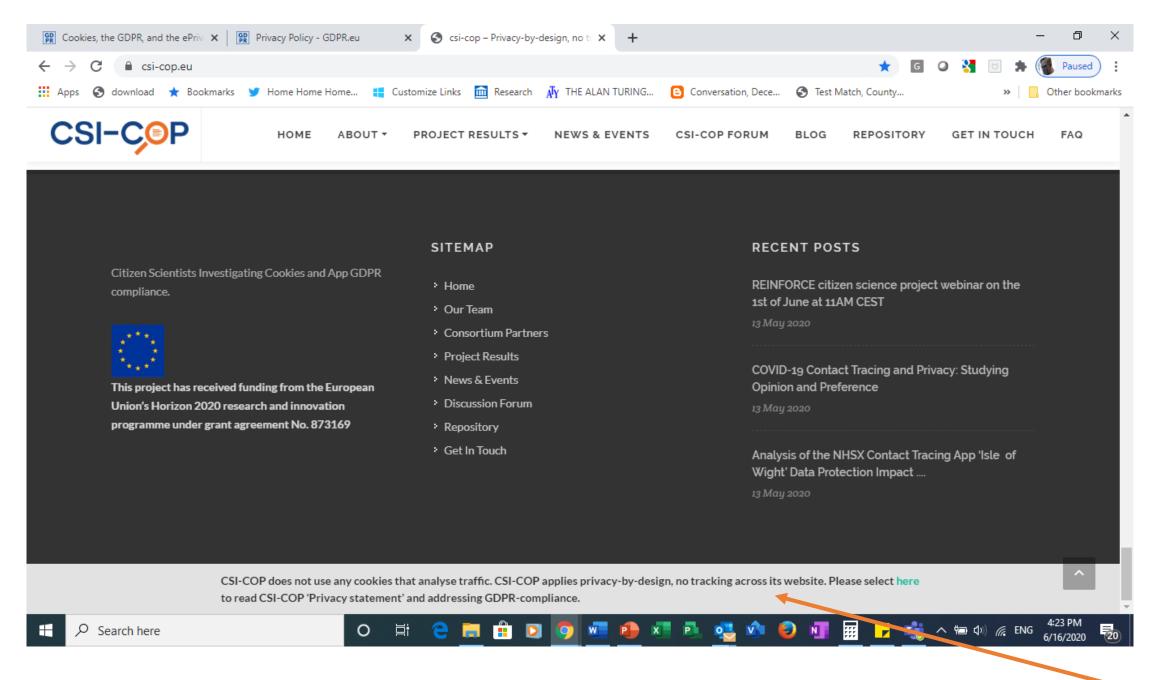
CSI-COP website

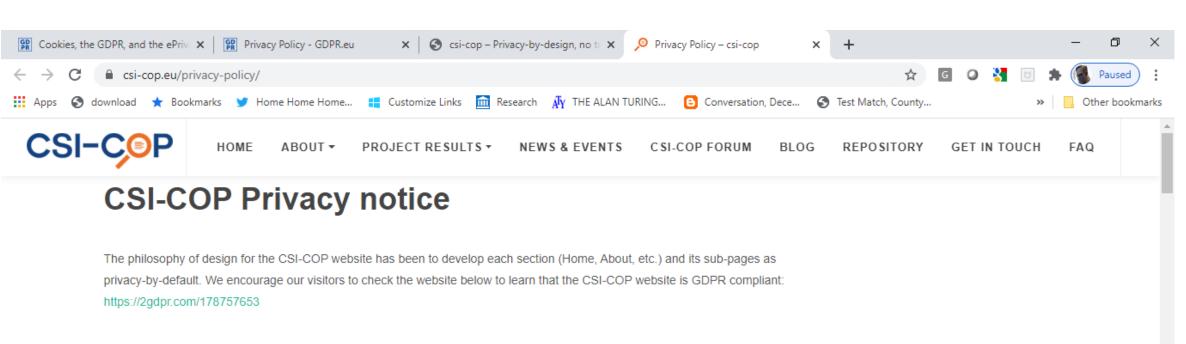
Philosophy:

- Privacy-by-design
- No-tracking website

• How:

- Reverse engineer WordPress to extract embedded cookies that track visitors
- Intention: trust between citizen scientists and other visitors to CSI-COP site





The development has involved ensuring the website is free of tracking visitors' movement and navigation around CSI-COP pages. This means the CSI-COP website does not track from which website a visitor arrived on to a CSI-COP web page, nor does CSI-COP website monitor or record the 'session' of each visitor. To this end the CSI-COP website has no visitor tracking on its web pages.

https://csi-cop.eu/privacy-policy/

Does this mean the CSI-COP website carries no cookies (small text files, please see FAQ page and further information below)? The site is forced to include a small number of cookies to function, although these do not result in a visitor being tracked.

Home Page

For example, on the Home page, there are 'third-party' requests from Google for use of its fonts. CSI-COP is the 'first-party' for the CSI-COP website. A 'third-party' means a party that is not related in any way with the first party, in this case, CSI-COP. The third-party requests relate to the fonts used on each of the pages on CSI-COP's website.

Why does CSI-COP use Google fonts? These were embedded in the WordPress template CSI-COP adopted to create its web



































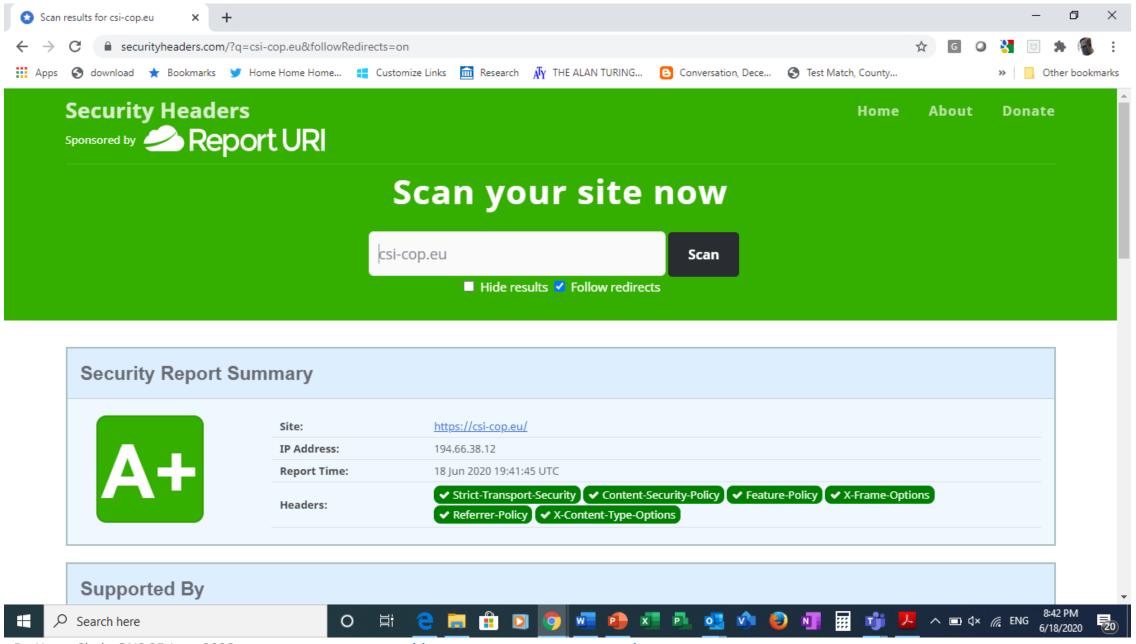


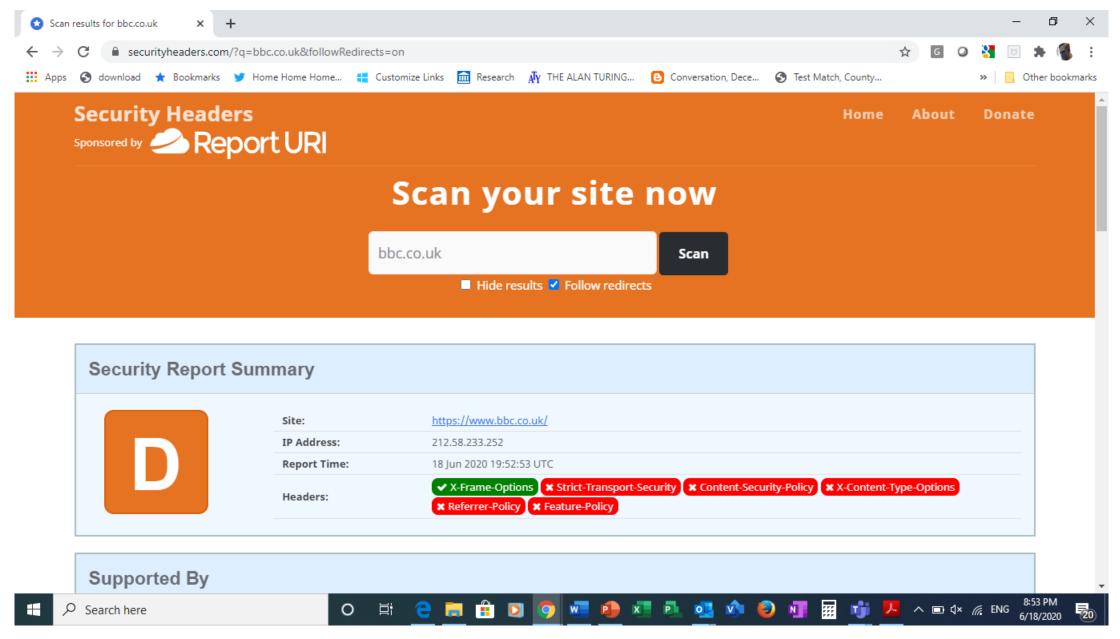




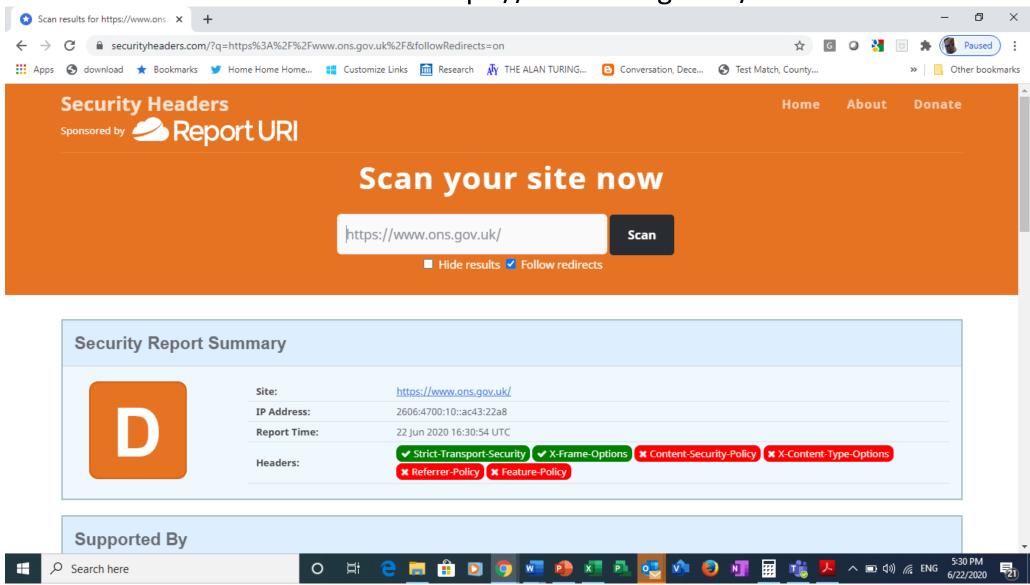








ONS website: https://www.ons.gov.uk/



Other main CSI-COP activities

- Citizen science- Stakeholder cafés: platform to discuss data protection and privacy issues in digital technologies
- Parent-Teacher Round-tables with invited CSI-COP citizen scientists explaining their experiences on discovering trackers embedded in website cookies and apps.
- Pro-privacy champions arising from the citizen scientists in CSI-COP connected to the university partners and engaged for Outreach events promoting data protection (project legacy)
- Working group created from the partners to maintain and update the repository for maximum use to better inform the public on digital tracking and reduce the problem of tracking-by-default through advocating pro-privacy software development

Indicators to Measure CSI-COP Impact

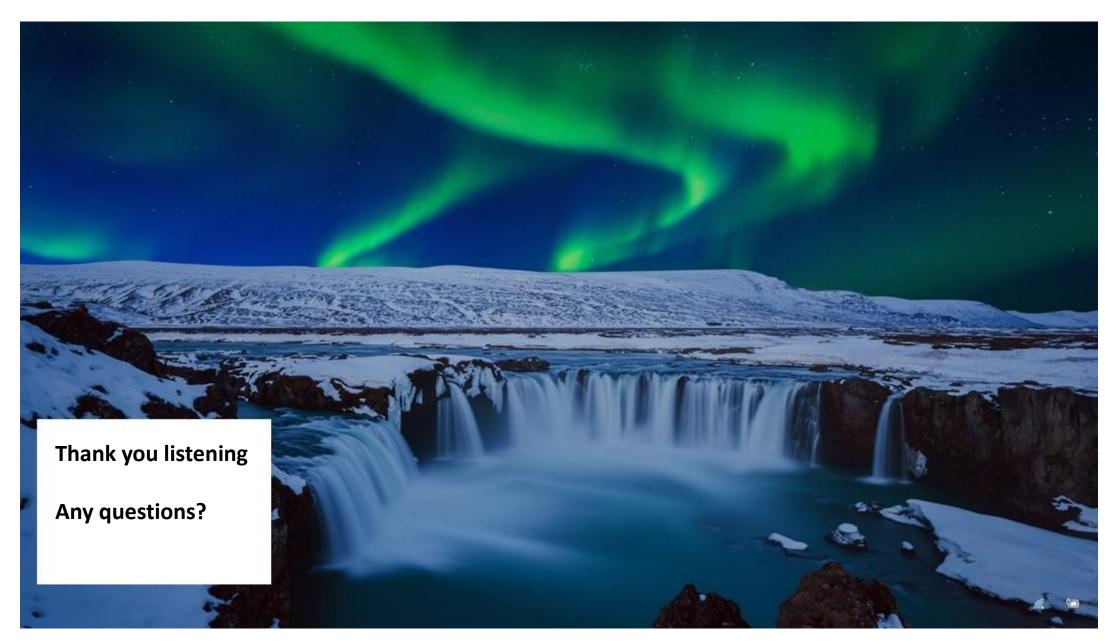
- Impact measures from MoRRI &UNSDG indicators include:
 - **Gender equality** (GE): Balanced recruitment of male, female and other gender citizen scientists
 - Science Literacy and education (SLSE): Responsible Research and Innovation (RRI) Training; Science communication through informal education into GDPR
 - Ethics (E) Ethics at the level of the University and other partners in CSI-COP consortium
 - Public Engagement (PE): various
 - Open Access (OA): Open Access to research and to CSI-COP's Innovation
 - Partnership for Goals: Encourage and promote effective public, public-private and civil society partnerships

Expected Impacts

- Impact on Responsible Research and Innovation (RRI)
- Development of new knowledge and innovations by citizen scientists
- Impact on the science: GDPR compliance
- Other impacts include:
 - Screen time
 - Staying safer online
 - Upskilling the public and educators

References

- CSI-COP EU Cordis page: https://cordis.europa.eu/project/id/873169
- CSI-COP website: https://www.csi-cop.eu
- Manjoo, F. (2020). 'I Visited 47 Sites. Hundreds of Trackers Followed Me'. New York Times. Accessed June 1, 2020 from here: https://www.nytimes.com/interactive/2019/08/23/opinion/data-internet-privacy-tracking.html
- Kelion, L. (2020). Babylon Health admits GP app suffered a data breach. BBC Technology, 9 June, 2020: https://www.bbc.co.uk/news/technology-52986629
- Powles, J. and Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. Health and Technologies, 7 (351-367). DOI: 10.1007/s12553-017-0179-1
- SciStarter: https://scistarter.org/
- Syel, R. (2018). Matt Hancock accused of breaching code over GP app endorsement. *Guardian News*, 30 Nov 2018. Accessible from here: https://www.theguardian.com/politics/2018/nov/30/matt-hancock-accused-of-breaching-code-over-gp-app-endorsement



Dr. Huma Shah, ONS 25 June 2020