Exam Revision 2

INFR11158/11230 Usable Security and Privacy

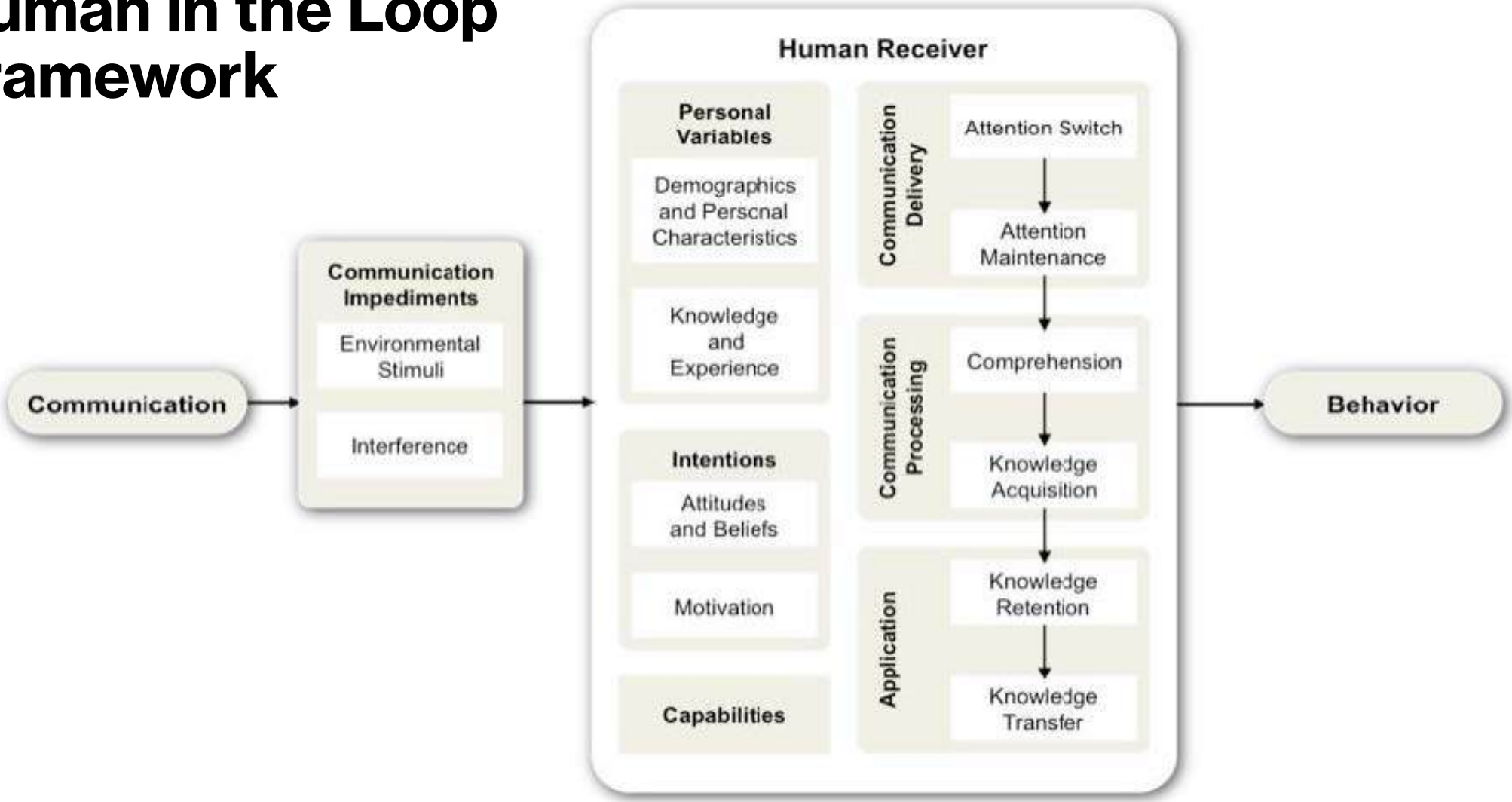
Dr. Jingjie Li

04/04/2025

THE UNIVERSITY of EDINBURGH

Framework and Topics

Human in the Loop Framework



https://medium.com/@ezgineer/usable-security-and-privacy-introduction-d676abc8c61d

Other Frameworks: What are they used for, and how to use them?

- NEAT
- SPRUCE
- Privacy by design
- Contextual integrity
- ....

A TAXONOMY OF PRIVACY

INFORMATION PROCESSING

AGGREGATION
Combining of various pieces of personal information

A credit bureau combining an individual's payment history from multiple creditors

SECONDARY USE
Using personal information for a purpose other than the purpose for which it was collected

The U.S. Government using census data collected for the purpose of apportioning Congressional districts to identify and intern those of Japanese descent in WWII

EXCLUSION
Failing to let an individual know about the information that others have about them and participate in its handling or use

A company using customer call history, without the customer's knowledge, to shift their order in a queue (i.e. "Your call will be answered in the order [NOT] received")

INSECURITY
Failing to protect information

An ecommerce website allowing others to view an individual's purchase history by changing the URL (e.g. enterprivacy.com?id=123)

IDENTIFICATION
Linking of information to an individual. [Sometimes called 'singling out']

A researcher linking medical files to the Governor of a state using only date of birth, zip code and gender.

COLLECTION

SURVEILLANCE
Watching, listening to, or recording of a person's activities

A website monitoring cursor movements of a visitor while visiting the website

INTERROGATION
Questioning or probing for personal information

An interviewer asking an inappropriate question, such as marital status, during an employment interview.

INVASION

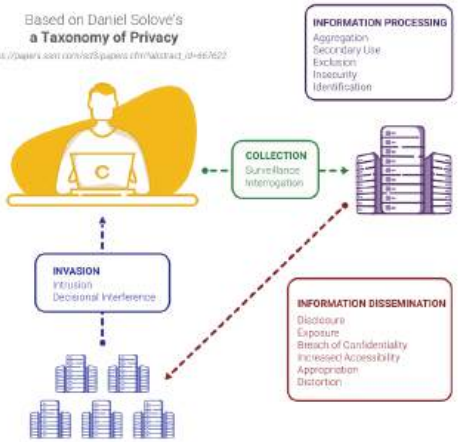
INTRUSION
Disturbing a person's tranquility or solitude

An augmented reality game directing players onto private residential property

DECISIONAL INTERFERENCE
Intruding into a person's decision making regarding their private affairs

A payment processor declining transactions for contraceptives.

Based on Daniel Solove's
a Taxonomy of Privacy



INFORMATION DISSEMINATION

DISCLOSURE
Revealing truthful information about a person that impacts their security or the way others judge their character

A government agency revealing an individual's address to a stalker, resulting in the individual's murder.

EXPOSURE
Revealing a person's nudity, grief, or bodily functions

A store forcing a customer to remove clothing revealing a colostomy bag

BREACH OF CONFIDENTIALITY
Breaking a promise to keep a person's information confidential.

A doctor revealing patient information to friends on a social media website.

INCREASED ACCESSIBILITY
Amplifying the accessibility of personal information

A court making proceeding searchable on the internet without redacting personal information

APPROPRIATION
Using an individual's identity to serve the aims and interests of another

A social media site using customer's images in advertising.

DISTORTION
Disseminating false or misleading information about a person

A creditor reporting a paid bill as unpaid to a credit bureau.

PRIVACY BY DESIGN

Version 6 (2022)

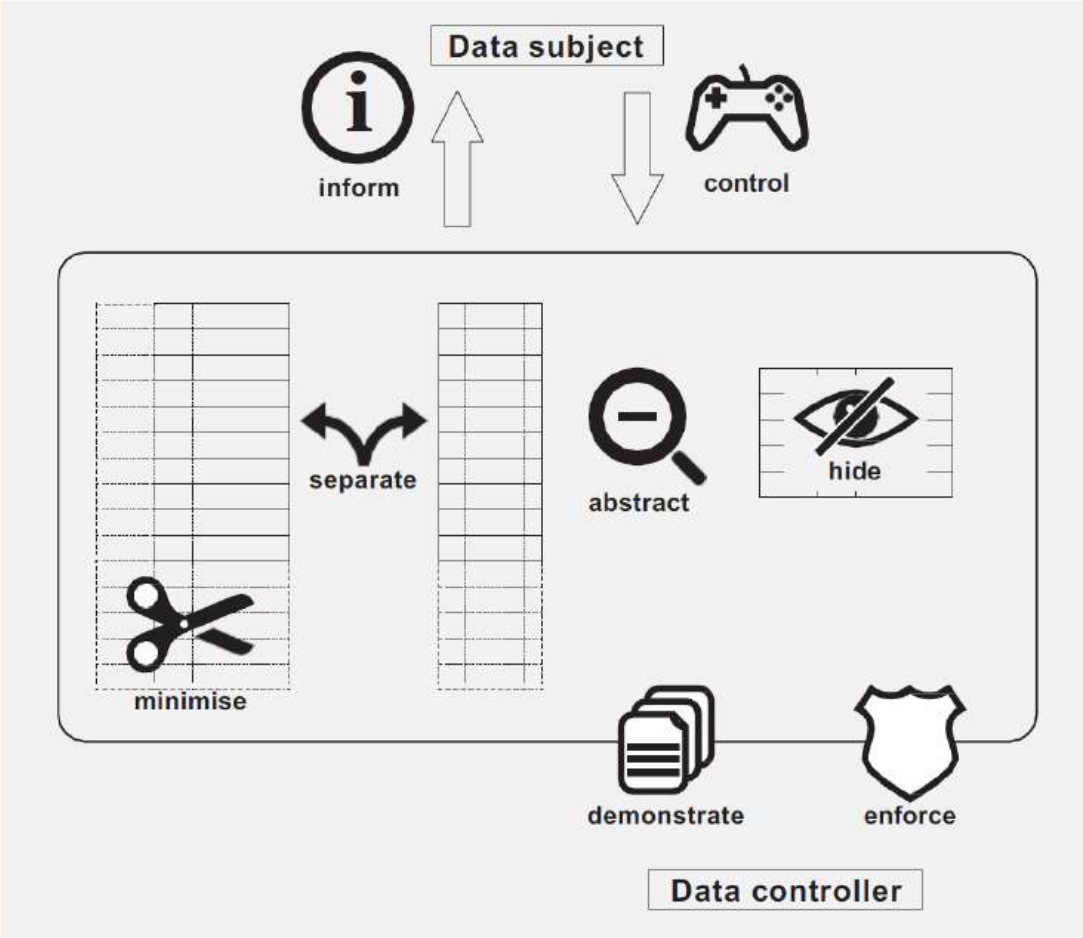
https://privacybydesign.training

Privacy by design – definition

Framework for building privacy proactively into new systems, proposed in 2009.Widely accepted as an international standard for good privacy engineering. GDPR also basis some of its principles on Privacy by Design.

- **Proactive** not Reactive; **Preventative** not Remedial
- Privacy as the **Default**
- Privacy **Embedded** into Design
- **Full** Functionality – Positive-Sum, not Zero-Sum
- End-to-End Security – Lifecycle Protection
- Visibility and Transparency
- **Respect** for User Privacy

Privacy by design – strategies



NEAT

Necessary – Can you change the architecture to eliminate or defer this user decision? Interrupt users only when necessary.

Explained - Does your user experience present all the information the user needs to make this decision? Explain the decision users need to make with information **(See SPRUCE)**

Actionable – Have you determined a set of steps the user will realistically be able to take to make the decision correctly? Give steps in all scenarios (e.g., benign vs malicious)

Tested – Have you checked that your user experience is NEAT for all scenarios, both benign and malicious? Have you tested it on a human who is not a member of your team? Do usability testing.

SPRUCE

Source – State who or what is asking the user to make a decision

Process – Give the user actionable steps to follow to make a good decision

Risk – Explain what bad thing could happen if they user makes the wrong decision

Unique – Knowledge the user has – Tell the user what information they bring to the decision regarding the context

Choices – List available options and clearly recommend one

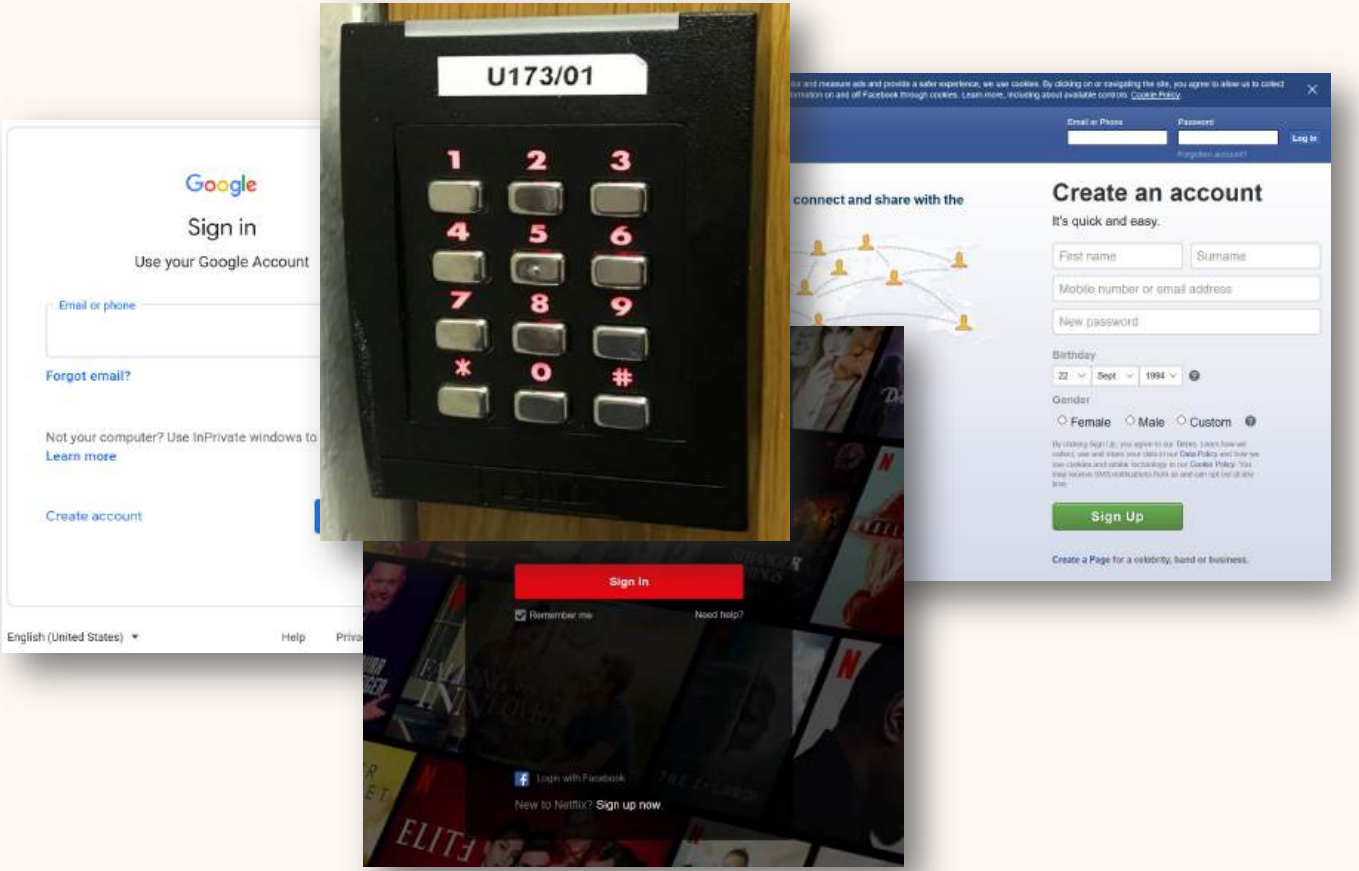
Evidence – Highlight information the user should factor in or exclude in making a decision

Privacy space framework

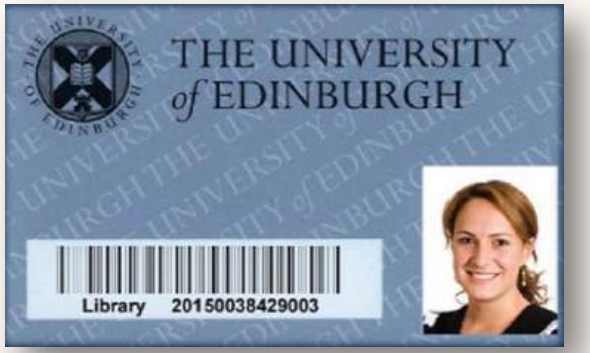
Category	Description	Examples
Awareness	Informative	Display information about trackers on current webpage, whether location is being sent
Detection	Actively look for problems	Find trackers on current webpage
Prevention	Used as a precaution	Encryption tools, anonymity tools
Response	Taking action after a problem is detected	Tracking blocker
Recovery	Help you get back to normal	Patching bugs

Benjamin Brunk. A user-centric privacy space framework. In Cranor and Gafinkel, eds. *Security and Usability*. O'Reilly 2005. p. 401-420.

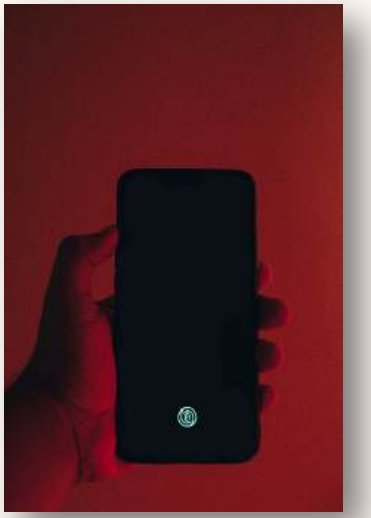
Authentication



What you know



What you have



Who you are

A good authentication method:

User friendly

- Memory effortless
- Scalable for users
- Nothing to carry
- Physically effortless
- Easy to learn
- Efficient to use
- Infrequent errors
- Easy to recover from loss

Reasonable to implement

- Accessible
- Negligible cost per user
- Server compatible
- Browser compatible
- Mature
- Non-proprietary

Protects against attacks

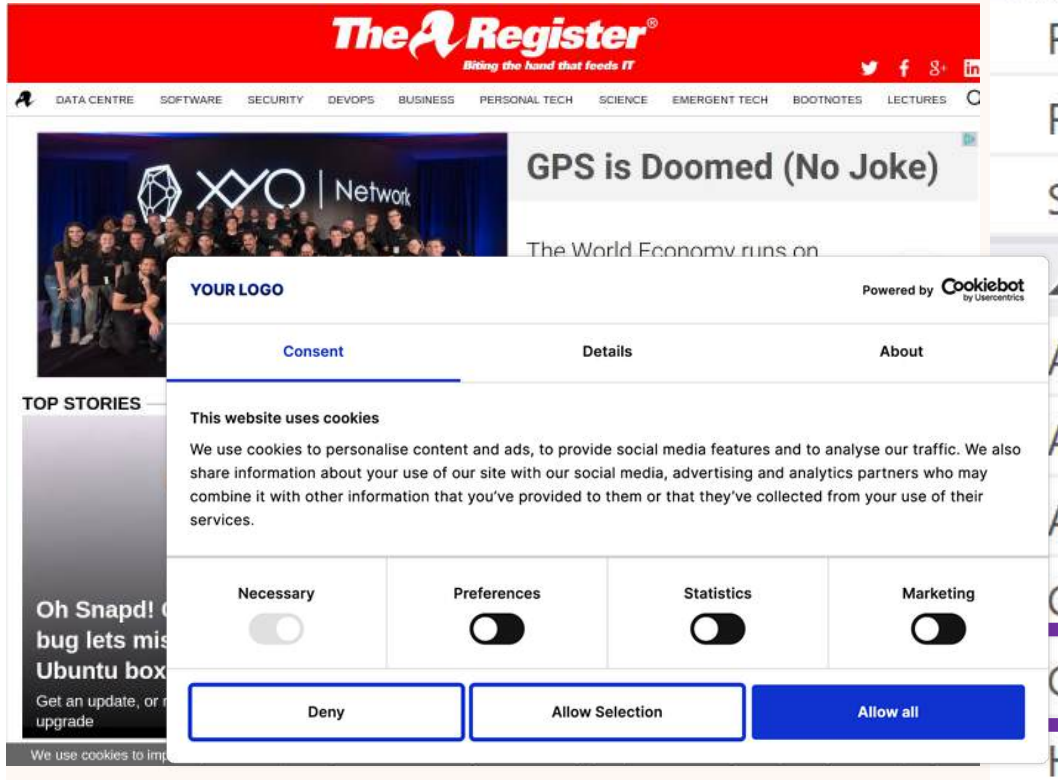
- Resilient to:
  - Physical observation
  - Targeted impersonation
  - Throttled guessing
  - Unthrottled guessing
  - Internal observation
  - Leaks from other verifiers
  - Phishing
  - Theft
- No trusted third party
- Requiring explicit consent
- Unlinkable

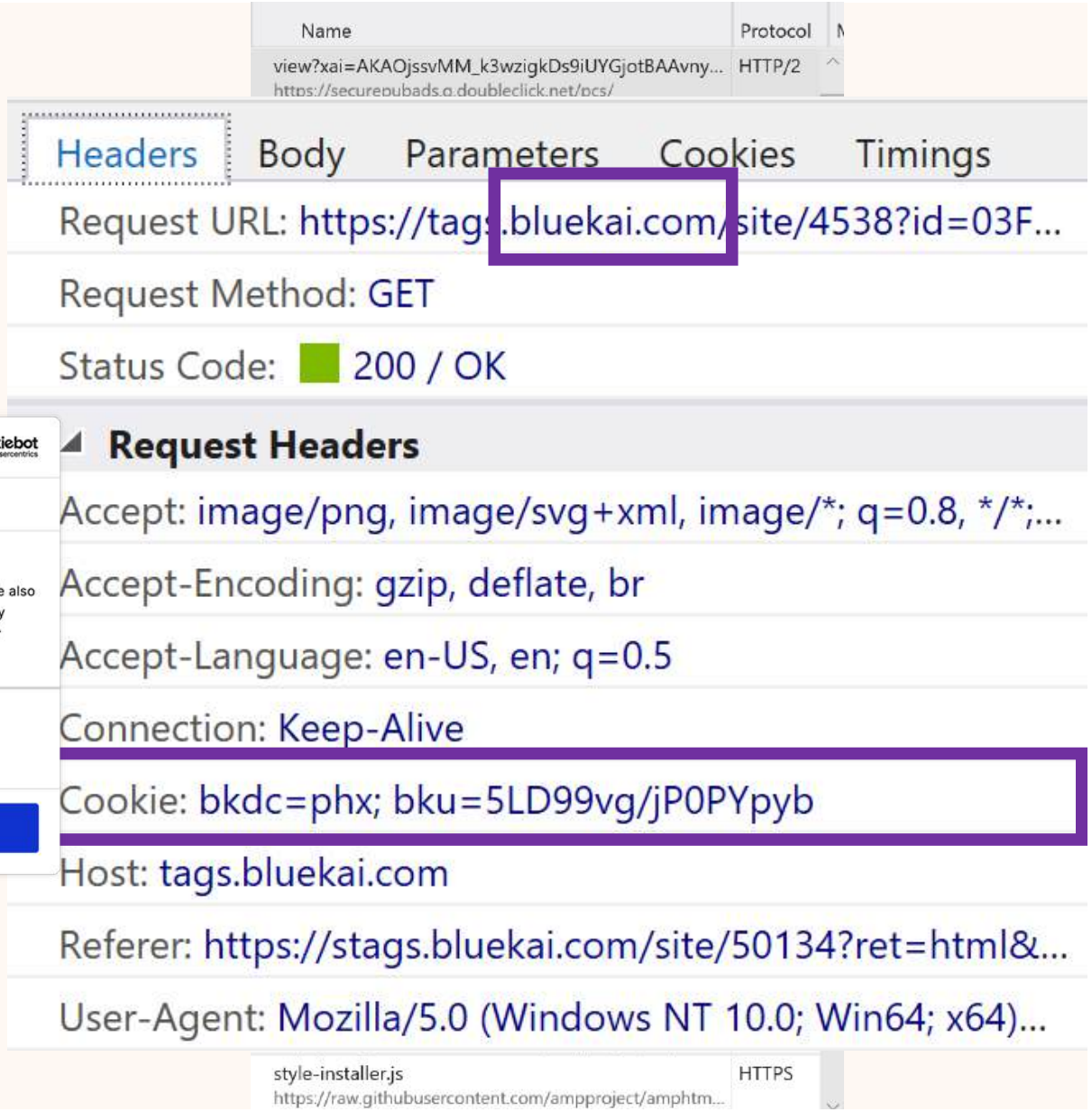
Attributes of a “good” biometric feature

1. **Universality:** Does everyone have it?
2. **Distinctiveness:** Is it different for everyone?
3. **Permanence:** Does the feature change over time/age?
   - bad: face, good: fingerprint
4. **Collectability:** How easy it is to collect/measure the feature?
   - Very hard: DNA, relatively easy: fingerprint
5. **Performance:** How difficult to match?
6. **Acceptability**
7. **Circumvention:** How easy to spoof?
   - Voice recognition

https://www.cse.msu.edu/~rossarun/pubs/RossBioIntro_CSVT2004.pdf

Cookie





https://ico.org.uk/for-organisations/direct-marketing-and-privacy-and-electronic-communications/guide-to-pecr/cookies-and-similar-technologies/

BEFORE OPT-OUT

Headers | Body | Parameters | Cookies | Timings

Request URL: https://tags.bluekai.com/site/4538?id=03F...

Request Method: GET

Status Code: 200 / OK

Request Headers

Accept: image/png, image/svg+xml, image/*; q=0.8, */*;...

Accept-Encoding: gzip, deflate, br

Accept-Language: en-US, en; q=0.5

Connection: Keep-Alive

Cookie: bkdc=phx; bku=5LD99vg/jP0PYpyb

Host: tags.bluekai.com

Referer: https://stags.bluekai.com/site/50134?ret=html&...

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64)...

AFTER OPT-OUT

Headers | Body | Parameters | Cookies | Timings

Request URL: https://stags.bluekai.com/site/50134?ret=h...

Request Method: GET

Status Code: 200 / OK

Request Headers

Accept: text/html, application/xhtml+xml, application/x...

Accept-Encoding: gzip, deflate, br

Accept-Language: en-US, en; q=0.5

Connection: Keep-Alive

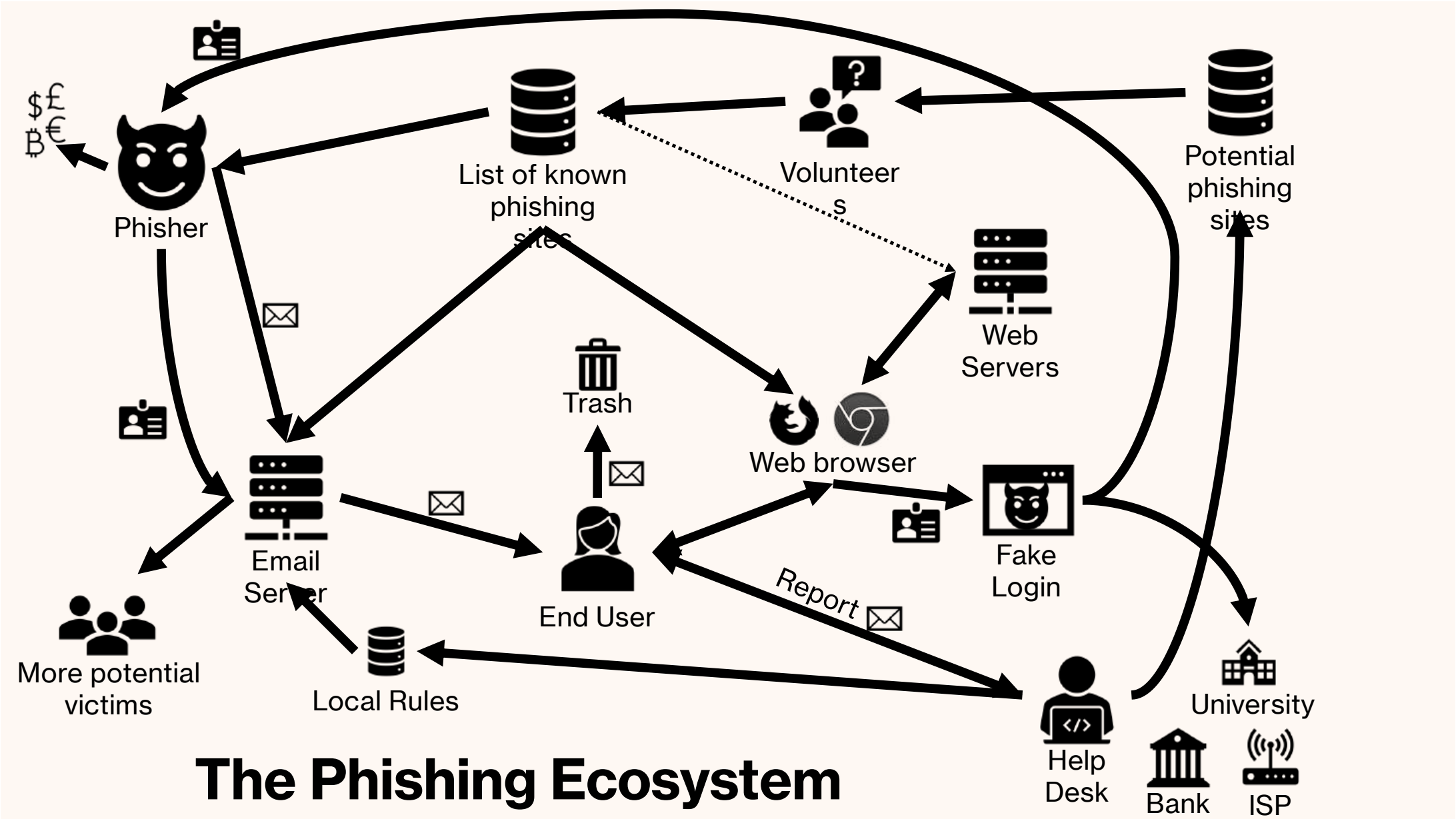
Cookie: bku=0000000000000000; BKIgnore=1; bkdc=phx

Host: stags.bluekai.com

Referer: https://www.nytimes.com/

Upgrade-Insecure-Requests: 1

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64)...


Phisher
List of known phishing sites
Volunteers
Potential phishing sites
Web Servers
Trash
Web browser
Email Server
End User
Fake Login
Report
More potential victims
Local Rules
Help Desk
University
Bank
ISP
The Phishing Ecosystem

Common phishing elements

- **Automated** – Typically directed against many people.
- **Impersonation** – Communication claims to be from someone trusted or that they are not. For example, from a bank.
- **Direction to a website** – Links that look like they go somewhere legitimate but in fact go somewhere controlled by the attacker.
- **Contain an attachment** – Attachment asks for information to be sent back or contains malicious code.
- **Authentication info requested** – The communication aims to get authentication information.

Main “solutions” against phishing

- Automatically block attacks using filters
- Train users
- Support users
- Improve protection of authentication credentials

Overview of Stanford Fraud Taxonomy

- Consumer Investment Fraud
  - Securities fraud
    - Equity investment fraud
      - Penny stock fraud
      - ...
    - ...
  - ...
- Consumer Products and Services Fraud
  - ...
    - *Phishing websites/emails/calls*
- Employment Fraud
- Prize and Grant Fraud
- Phantom Debt Collection Fraud
- Charity Fraud
- Relationship and Trust Fraud

All sorts of things need to be communicated to users

- **Questions** – “did you log in from this location?”
- **Warnings** – “the website has malicious software”
- **UI passive indicators** – the lock icon on the browser
- **UI active indicators** – “You need to generate a key”
- **Task-relevant information** – “Passwords should be 8 characters long and must have a capital letter.”
- **Educational** – “10 security behaviors you should do to protect yourself online”
- **Awareness** – “This phishing email has been going around, don’t fall for it.”

https://googleonlinesecurity.blogspot.com.au/2015/07/new-research-comparing-how-security.html

Access Control Matrix

Objects (files)

Subjects (users)

	a	b	c	d	e
jingjie	r,w	-	r,w, own	-	r
bob	-	-	r	r	r,w
alice	w, own	r	r	-	-
eve	r	r,w	r,w	-	r

Permitted operations

[Lampson, Graham, Denning; 1971]

Could be a very huge table to store and access!

ACL vs. Capabilities

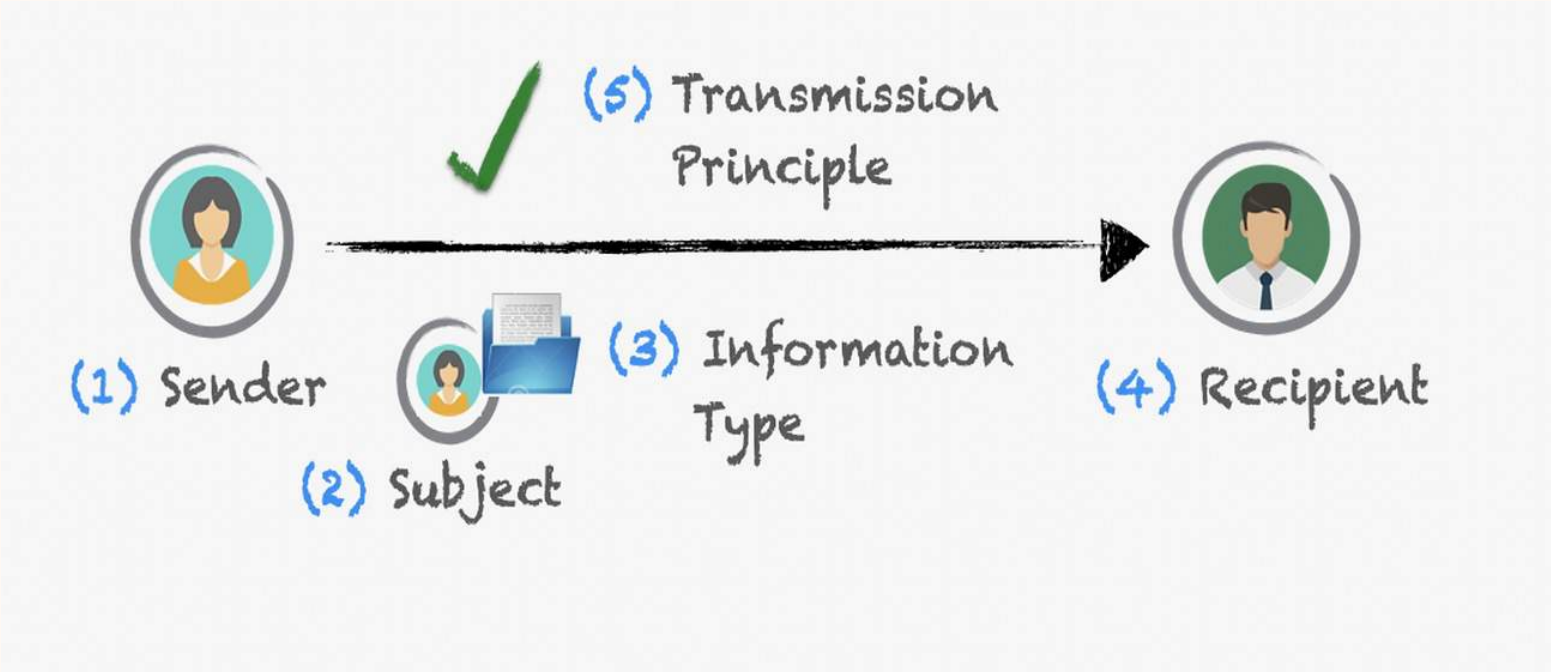
ACL

- Each file contains lists of user ids with their permissions (column in AC matrix)
- Check user/group against ACL
- Relies on authentication
- Inefficient run-time security checking

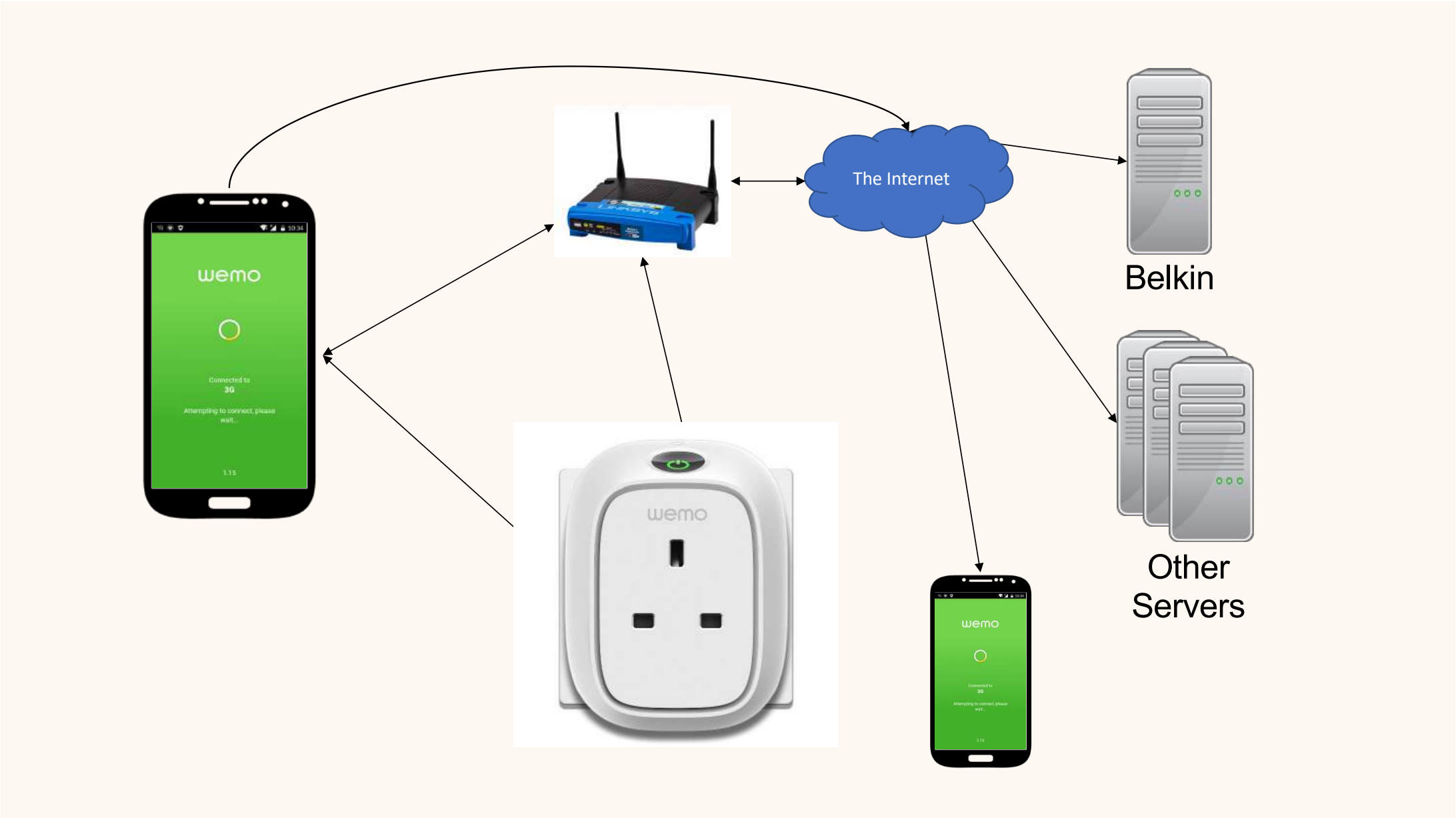
Capabilities

- Stores each user's capabilities (row in AC matrix)
- Check validity of capability
- Can be easily passed to other subjects (delegation)
- Hard to change a file's status globally, e.g., revocation

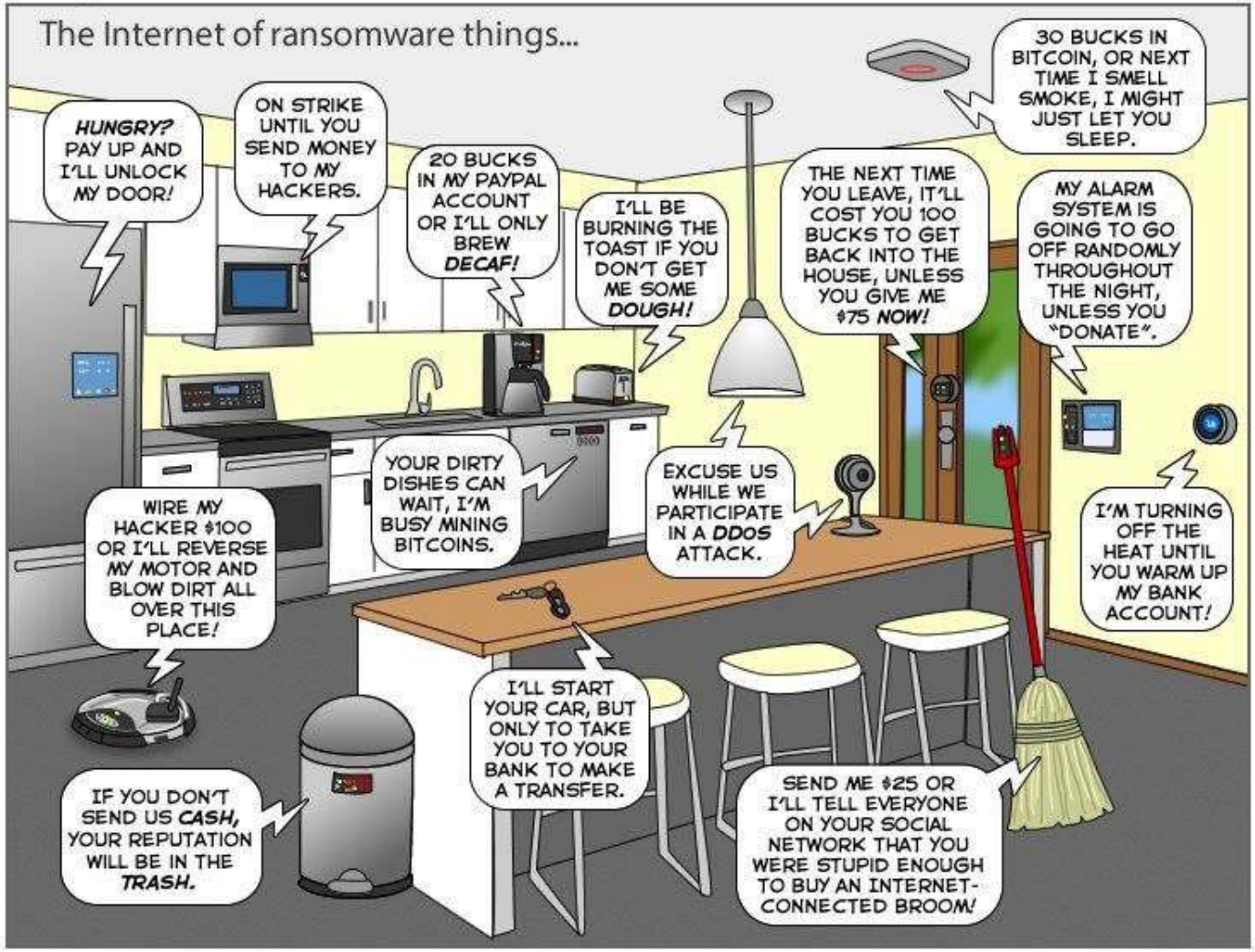
Contextual integrity



https://www.dli.tech.cornell.edu/post/privacy-policies-as-contextual-integrity-beyond-rules-compliance


wemo
Connected to
3G
Attempting to connect, please wait...
1.15
The Internet
Belkin
Other Servers
wemo
wemo

The Joy of Tech™ by Nitrozac & Snaggy



You can help us keep the comics coming by becoming a patron!
www.patreon/joyoftech

joyoftech.com

Adversarial Examples

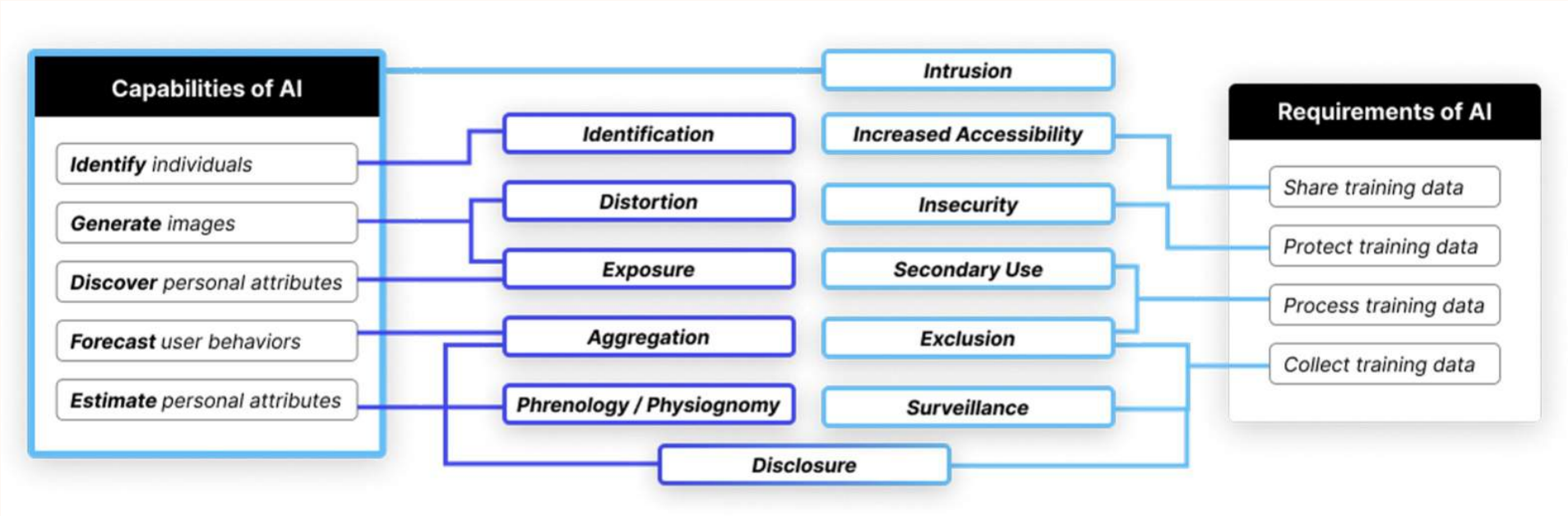
- Definition
  - Inputs to machine learning models that an attacker has intentionally designed to cause the model to make a mistake.
- Impact
  - Leads to incorrect AI decisions or misclassifications that seem correct to human operators.
- Methodology
  - Creating input samples that are slightly altered but cause significant errors in AI outputs.
  - Exploiting model vulnerabilities that are not easily detectable by humans.
- Countermeasures
  - Employing adversarial training methods.
  - Regularly updating and testing models against known adversarial attack techniques.

Prompt Injection

- Definition
  - Manipulation of AI's response by altering the input prompt or commands it receives.
- Impact
  - Can cause AI to produce undesired, biased, or harmful outputs.
- Methodology
  - Craft malicious input prompts to mislead AI.
  - Inject misleading context or information into the AI's operational environment.
- Countermeasures
  - Robust input validation and sanitization.
  - Implementation of authentication protocols to verify source integrity.

Data Poisoning

- Definition
  - Introducing malicious data into the AI's training set to corrupt its learning process.
- Impact
  - Results in a corrupted model that makes errors or biased decisions.
- Methodology
  - Insertion of subtly incorrect or biased data points into the training dataset.
  - Targeted manipulation to influence specific AI behaviors or outcomes.
- Countermeasures
  - Regular audits of training data.
  - Use of anomaly detection techniques to identify and remove corrupted data.


Capabilities of AI
Identify individuals
Generate images
Discover personal attributes
Forecast user behaviors
Estimate personal attributes
Intrusion
Identification
Increased Accessibility
Distortion
Insecurity
Exposure
Secondary Use
Aggregation
Exclusion
Phrenology / Physiognomy
Surveillance
Disclosure
Requirements of AI
Share training data
Protect training data
Process training data
Collect training data

Bug Bounty Stakeholders

- Bug hunter
- Platform
  - Operator
  - Triager
  - Mediator
- Vendor/Program
  - Reviewer/Security team
  - Developer
- End user

The Belmont Report (1974)

- Respect for persons
  - Protecting the autonomy of all people and treating them with courtesy and respect and allowing for informed consent. Researchers must be truthful and conduct no deception
- Beneficence
  - The philosophy of "Do no harm" while maximizing benefits for the research project and minimizing risks to the research subjects
- Justice
  - Ensuring reasonable, non-exploitative, and well-considered procedures are administered fairly – the fair distribution of costs and benefits to *potential* research participants – and equally.

http://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

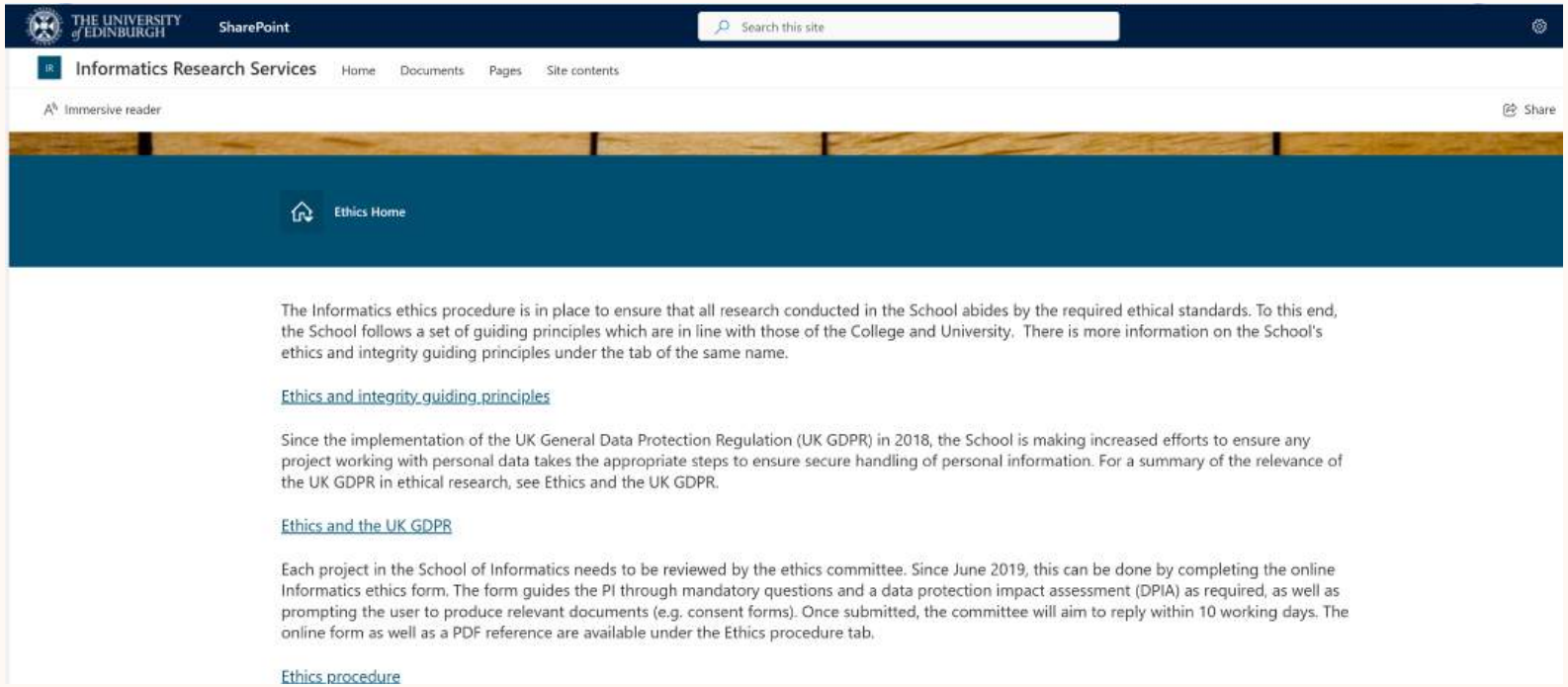
The Menlo Report (2012)

Principle	Application
Respect for Persons	Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection.
Beneficence	Do not harm; Maximize probable benefits and minimize probable harms; Systematically assess both risk of harm and benefit.
Justice	Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit; Selection of subjects should be fair, and burdens should be allocated equitably across impacted subjects.
Respect for Law and Public Interest	*Engage in legal due diligence; Be transparent in methods and results; Be accountable for actions.*

Consent in General Data Protection Regulation

The basic requirements for the effectiveness of a valid legal consent are defined in Article 7 and specified further in recital 32 of the GDPR. **Consent** must be freely given, specific, **informed** and unambiguous. In order to obtain freely given consent, it must be given on a voluntary basis. The element “free” implies a real choice by the data subject….

1998 Act:	GDPR:
Principle 1 – fair and lawful	Principle (a) – lawfulness, fairness and transparency
Principle 2 – purposes	Principle (b) – purpose limitation
Principle 3 – adequacy	Principle (c) – data minimisation
Principle 4 – accuracy	Principle (d) – accuracy
Principle 5 - retention	Principle (e) – storage limitation
Principle 6 – rights	No principle – separate provisions in Chapter III
Principle 7 – security	Principle (f) – integrity and confidentiality
Principle 8 – international transfers	No principle – separate provisions in Chapter V
(no equivalent)	Accountability principle



THE UNIVERSITY of EDINBURGH SharePoint

Search this site

Informatics Research Services Home Documents Pages Site contents

Immersive reader

Share

Ethics Home

The Informatics ethics procedure is in place to ensure that all research conducted in the School abides by the required ethical standards. To this end, the School follows a set of guiding principles which are in line with those of the College and University. There is more information on the School's ethics and integrity guiding principles under the tab of the same name.

Ethics and integrity guiding principles

Since the implementation of the UK General Data Protection Regulation (UK GDPR) in 2018, the School is making increased efforts to ensure any project working with personal data takes the appropriate steps to ensure secure handling of personal information. For a summary of the relevance of the UK GDPR in ethical research, see Ethics and the UK GDPR.

Ethics and the UK GDPR

Each project in the School of Informatics needs to be reviewed by the ethics committee. Since June 2019, this can be done by completing the online Informatics ethics form. The form guides the PI through mandatory questions and a data protection impact assessment (DPIA) as required, as well as prompting the user to produce relevant documents (e.g. consent forms). Once submitted, the committee will aim to reply within 10 working days. The online form as well as a PDF reference are available under the Ethics procedure tab.

Ethics procedure

https://uoe.sharepoint.com/sites/inf-researchservices/SitePages/Ethics-and-integrity.aspx

Some ethical practices for social media research

- Follow the terms of use
- Obtain informed consent when possible
- Check our ethics guidelines for more! https://resource.ppls.ed.ac.uk/lelethics/index.php/frequently-asked-questions/research-with-social-media-data/

Some examples of at-risk groups

"We define a user(s) as being at-risk if they face an elevated likelihood of an attack to their digital safety, have factors that influence or exacerbate their chances of being targeted, and/or experience heightened harm as a result of a digitally-mediated attack"

- Survivors of intimate partner violence
- Political activist
- Identity based marginalization (e.g., queer, women, people of color....)

Safe practices for at risk users

Category	ID	Digital-safety practices	Example papers
Professional partnerships & Ethical review	SP1	Elicit expert (academic) opinion on topic area	[17, 31, 67, 70, 82, 83, 112, 132, 136]
	SP2	Form professional partnerships (e.g., support services for at-risk users)	[44, 52, 72, 80, 82, 99, 105, 124, 134, 145]
	SP3	Invite and include an at-risk user to join research team	[17, 83, 97, 112]
	SP4	Seek external (non-institutional) ethical review approval or monitoring	[30, 43, 44, 78]
Positionality & Participant engagement	SP5	Build rapport with participants for understanding digital-safety needs	[1, 33, 34, 38, 73, 91, 97, 113, 137]
	SP6	Conduct pilot studies with general (non-at-risk) users	[5, 30, 33, 64, 67, 95, 101]
	SP7	Conduct studies with proxies for at-risk users (e.g., advocacy groups)	[2, 24, 33, 70, 74, 104, 132]
	SP8	Include researchers whose identities affirm participants'	[2, 6, 38, 64, 97, 110, 112, 113, 132, 134]
	SP9	Practice responsiveness in data collection sessions to potential threats	[3, 38, 49, 89, 100, 101, 124, 127, 128, 132]
	SP10	Provide professional therapeutic support for emotive topics	[7, 11, 30, 48, 95, 100, 101, 115, 144]
	SP11	Train team members in working with digital-safety risks	[7, 38, 115, 121]
Privacy-preserving data collection	SP12	Discourage participant self-disclosure (e.g., personal histories)	[1, 7, 25, 52, 70, 75, 118, 123, 137, 144]
	SP13	Focus data collection on supporting participant safety needs	[24, 34, 38, 66, 81, 97, 120, 121, 123, 129]
	SP14	Do not collect or ask for participant demographic data	[17, 26, 64, 83, 84, 104, 120, 124, 136, 145]
	SP15	Do not collect personally identifiable information on participants	[30, 43, 44, 52, 54, 58, 73, 85, 95, 143]
	SP16	Implement protocols for researchers to prevent stalking by adversaries	[30, 60, 80]
	SP17	Separate potential threats from at-risk users during data collection	[6, 72, 88, 96, 97, 100, 110, 115]
	SP18	Permit participants to contribute false information (e.g., pseudonyms)	[17, 54, 58, 78, 83, 100]
	SP19	Offer participants many modalities to contribute (e.g., audio, notes)	[4, 7, 24, 34, 57, 67, 90, 107, 117, 130]
	SP20	Secure confidentiality and privacy of online and in-person research sites	[6, 24, 30, 43, 44, 77, 100, 113, 134, 139]
Secure data storage & processing	SP21	Implement strict data access control measures for research data	[1, 7, 34, 51, 80, 112, 134, 136, 139, 147]
	SP22	Redact participant information prior to analysis by research team	[59, 86, 95, 107, 114, 128, 130, 140, 143, 156]
	SP23	Use encryption for research data in-transit and at-rest	[52, 60, 75, 85, 86, 87, 101]
	SP24	Use non-encrypted safe storage for research data in-transit and at-rest	[7, 30, 34, 90, 97, 114, 130, 132]
Researcher accountability	SP25	Conduct data collection sessions around participant schedules	[1, 35, 54, 65, 97, 111, 120, 128, 139]
	SP26	Offer formal proof of identity as professional researchers	[70, 82, 97, 112, 114, 115]
	SP27	Only use data from publicly accessible sites (e.g., no authorization)	[11, 32, 40, 97, 103, 138, 147, 155]
	SP28	Provide proportional incentives to participants for contributions	[54, 64, 72, 73, 82, 110, 134, 139, 145, 151]
	SP29	Be transparent with participants about risks incurred by research	[24, 26, 38, 54, 57, 69, 95, 110, 113, 128]
Sharing & evaluating deliverables	SP30	Do not attribute reported data contributions with participant identifiers	[7, 8, 9, 34, 55, 84, 114, 117, 134]
	SP31	Do not report participant demographics in research deliverables	[17, 24, 43, 77, 78, 83, 117, 120, 144, 145]
	SP32	Do not report participant names, pseudonyms, or identifiers	[9, 48, 71, 78, 101, 114, 121, 143, 145, 155]
	SP33	Paraphrase or withhold sources of data (e.g., websites they use)	[2, 9, 17, 40, 59, 69, 78, 123, 136, 155]
	SP34	Evaluate research deliverables for adversarial feedback or education	[34, 38, 44, 59, 82, 113]
	SP35	Selectively edit participant data in research deliverables	[7, 9, 11, 40, 55, 124, 139, 140, 150, 151]
	SP36	Provide participants control of their contributions (e.g., permit redaction)	[7, 47, 54, 75, 91, 113, 114, 117, 136]

Safe practices for at risk users

ID	Strategy title	Description	Example digital-safety practices
S1	Engage experts early	Consult or partner with domain experts from the beginning to inform and help facilitate safe research plans.	SP1, SP2, SP3, SP4, SP10
S2	Assess and mitigate risks by threat modeling	Apply the S&P practice of threat modeling to research protocols, and continuously update threat models to guide ongoing safety mitigations.	SP11, SP16, SP17, SP20
S3	Select the lowest risk method that addresses the research goals	Before soliciting at-risk users for high-touch methods like interviews, consider proxies (e.g., advocates), or indirect methods (e.g., online measurement).	SP6, SP7, SP12, SP14, SP15, SP27
S4	Respect that at-risk users self-manage risk	At-risk users are often experts in managing their safety risks. Give them choice in how they engage with research safety protocols, and respect the choices they make.	SP9, SP18, SP19, SP25, SP26, SP29
S5	Be an advocate for at-risk users' needs	Research, by its nature, can be extractive. Build reciprocity with at-risk users, and work to help them achieve their goals.	SP5, SP8, SP13, SP28, SP36
S6	Handle data and publications carefully	Data collection and analysis should follow security best-practice, and publications should avoid revealing identities or informing adversaries.	SP21, SP22, SP23, SP24, SP30, SP31, SP32, SP33, SP34, SP35