

The Ethics of Investigating Digital Worlds

Brian Ballsun-Stanton

INTS1302 Navigating Digital Worlds, Macquarie University

2026-08-11

Investigating means observing people

RESEARCH IS CONDUCTED WITH OR ABOUT PEOPLE

- An online space is a room with people in it.
- Investigating the space means watching the people who gather there.
- The National Statement is Australia's rulebook for research about people.
- Coursework sits below the review threshold, but the ethics still applies.

The National Statement:

Human research is research conducted with or about people, or their data or their biospecimens.

([NHMRC et al., 2025, p. 6](#))

Kaufman, defending Harvard's release of student Facebook data:

Would you require that someone sitting in a public square, observing individuals and taking notes on their behavior, would have to ask those individuals' consent in advance? We have not accessed any information not otherwise available on Facebook. We have not interviewed anyone, nor asked them for any information, nor made information about them public...

(Kaufman 2008, comment on ([Zimmer, 2008](#)))

An online space is a room with people in it.

- the cohort's own rooms, a subreddit, a Discord, the group chat where screenshots land
- the room does not know it is being watched

Investigating the space means watching the people who gather there.

- why bother with ethics at all
- the answer is definitional, watching people is what the regulator calls research

The National Statement is Australia's rulebook for research about people.

- first meeting, the National Statement, written by the National Health and Medical Research Council, the NHMRC, the rules every Australian university researcher answers to
- Assessment 1 borrows its reasoning, the same weighing earns the top band

Coursework sits below the review threshold, but the ethics still applies.

- a real project goes before a committee, training exercises are the named exemption

The exemption in the Statement's own words:

"the research is conducted as part of an educational training program in which the research activity is for training purposes only and where any outcomes or documentation are for program use only"

The National Statement: “Human research is research conducted...”

- “or their data” is the load-bearing phrase, the diary, the post, the like all count
- operative, not decorative, it repeats word for word in the operative section

Kaufman: “Would you require that someone sitting in a public square...”

- first meeting, Jason Kaufman, Harvard sociologist, led the T3 project, Tastes, Ties and Time, four years of one college cohort’s Facebook profiles, about 1,700 students ([Lewis et al., 2008](#))
- his defence, public data, ordinary observation, no consent needed
- his defence and the definition cannot both hold, watching people is research about people
- the dataset did not stay anonymous for long

Risk is a weighing

SEVERITY DEPENDS ON WHO BEARS IT

- Risk is likelihood multiplied by severity.
- The same act can be safe for one person and dangerous for another.
- A risk is acceptable when a named benefit justifies it.

The National Statement:

Research is ethically acceptable only when its potential benefits justify any risks involved in the research.

([NHMRC et al., 2025, p. 17](#))

Ceglowski:

In 2007 LiveJournal is sold to a Russian company, and a few years later, to everyone’s surprise, homophobia is elevated to state ideology.

([Ceglowski, 2015, p. 3](#))

Kramer and colleagues, for the mood experiment:

We show, via a massive (N = 689,003) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness.

([Kramer et al., 2014, p. 1](#))

Boellstorff:

The principle of care arises in part from asymmetrical power relations and imbalance of benefit between investigator and investigated. The investigator generally gains far more than the informant, garnering benefits that translate to jobs, money, and professional recognition.

([Boellstorff et al., 2012, p. 2](#))

Risk is likelihood multiplied by severity.

- everyday talk blurs the two parts, the regulator splits them
- a tiny chance of catastrophe is still a serious risk

The regulator naming the blur:

“when expressions such as “small risk” or “high risk” are used, they usually refer (often ambiguously) both to the chance (probability) of experiencing a harm and the severity (magnitude) of the envisioned harm.” ([NHMRC et al., 2025, p. 15](#))

The same act can be safe for one person and dangerous for another.

- first meeting, Cegłowski, runs the Pinboard bookmarking site, a tech insider arguing against his own industry's data habits
- his case is a hypothetical built on real events

Cegłowski's setup:

"Closer to our time, consider the hypothetical case of a gay blogger in Moscow who opens a LiveJournal account in 2004, to keep a private diary." (Cegłowski, 2015, p. 3)

- in 2004 writing it is legal, the audience is chosen, the risk looks small
- the diary sits on a company's servers, storage outlives the moment of writing

A risk is acceptable when a named benefit justifies it.

- both pans filled, the diary gave him community and a voice
- collecting and storing forever is itself the serious decision

Cegłowski's weighing:

"So when you commit to collecting personal data, and storing it forever, you're making a serious decision." (Cegłowski, 2015, p. 3)

The National Statement: "Research is ethically acceptable only when..."

- the whole apparatus in one sentence, benefit weighed against risk

Cegłowski: "In 2007 LiveJournal is sold..."

- the man changed nothing, the ground moved under him
- hypothetical diarist, real timeline, the sale and the law both happened
- severity depends on who bears it, a straight diarist faced no new danger

Kramer and colleagues, for the mood experiment: "We show, via a massive..."

- first meeting, Kramer, Facebook's own data scientist, with academic co-authors at Cornell and UCSF, the 2014 mood experiment
- the experimenter's selling point is the subjects' unawareness, the asymmetry in one sentence
- the benefit lands with the platform, the risk lands on 689,003 people who never knew

Boellstorff: "The principle of care..."

- first meeting, Boellstorff, anthropologist of virtual worlds, wrote the field's ethnography handbook
- who gains and who bears, jobs and recognition on one side, the risk on the other
- who bears it and who benefits, stated as a principle, argued rather than measured

Every view is from somewhere

NOWHERE IS ALSO A POSITION

- Every observation happens from somewhere.

Haraway:

Relativism is a way of being nowhere while claiming to be everywhere equally.

- Claiming to see from nowhere is itself a position.
- The claim hides the people it cannot see.

([Haraway, 1988, p. 11](#))

Markham:

platform and service providers assert that their algorithms are impartial, functioning through rule-based calculations on objective data points without any “subjective” interference.

([Markham et al., 2018, p. 4](#))

Haraway:

*I am arguing for the view from a body, always a complex, contradictory, structuring, and structured body, versus the view from above, from nowhere, from simplicity. **Only the god trick is forbidden.***

([Haraway, 1988, p. 16](#))

Every observation happens from somewhere.

- logged out, Australian, adult, on a campus network, a different account sees a different platform
- somebody chose the categories before the data arrived

First meeting, boyd and Crawford, two data-studies scholars whose questions organised the field:

“Do numbers speak for themselves? We believe the answer is no.” ([Boyd & Crawford, 2012, p. 6](#))

Claiming to see from nowhere is itself a position.

- first meeting, Haraway, philosopher of science, 1988, named the god trick
- the god trick, the illusion of seeing everything from nowhere the way a god would, no body, no position, nobody to hold accountable
- platforms sell the same trick, impartial algorithms, objective data points

The claim hides the people it cannot see.

- private accounts, deleted posts, the people who left
- absent from the data, absent from the findings
- a view from a body can be argued with, a view from nowhere cannot

Haraway: “Relativism is a way of being nowhere...”

- relativism and the god trick are the same dodge, no position, no answerability

Markham: “platform and service providers assert...”

- first meeting, Markham, internet ethnographer, wrote the field’s method books and its ethics guidelines
- the god trick in a press release

Haraway: “I am arguing for the view from a body...”

- the bold sentence, only the god trick is forbidden
- situated does not mean subjective, a view from a body can answer for itself

Who bears it, and who benefits?

RECAP A

- Investigating an online space means observing people.

- Risk is likelihood times severity. Severity depends on who bears it.
- Every view is from somewhere.

Investigating an online space means observing people.

- the diarist was observed in a public room, he never chose his observers

Risk is likelihood times severity.

- the severity moved underneath him, the platform's sale moved it, not his choices

Every view is from somewhere.

- his room was seen from one particular somewhere, the buyer's somewhere was not his

Who bears it, and who benefits?

- the diarist bore it, LiveJournal and its buyer took the benefit

The rules were earned

EVERY RULE IS WRITTEN IN BLOOD

- Harm came first. The rules came after.
- Nuremberg 1947 answers the camp experiments. Belmont 1979 answers Tuskegee.
- Online research has its own scars. Harvard released 1,700 student profiles in 2008. Facebook ran its mood experiment on 689,003 users in 2014.
- The rules take decades to write. Online platforms have not existed that long.

The Nuremberg Code:

The voluntary consent of the human subject is absolutely essential.

([“The Nuremberg Code \(1947\),” 1996, p. 1](#))

The National Statement:

On rare occasions the practice of research has even involved the deliberate and appalling violation of human beings, notoriously, the Second World War experiments in detention and concentration camps.

([NHMRC et al., 2025, p. 6](#))

Facebook, after the mood experiment:

Although this subject matter was important to research, we were unprepared for the reaction the paper received when it was published and have taken to heart the comments and criticism. It is clear now that there are things we should have done differently.

([Schroepfer, 2014](#))

Harm came first. The rules came after.

- the sequence is documented, not rhetorical

First meeting, the Belmont Report, America's founding research ethics document, its account of the first code:

“During the Nuremberg War Crime Trials, the Nuremberg code was drafted as a set of standards for judging physicians and scientists who had conducted biomedical experiments on concentration camp prisoners.” ([National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 3](#))

Nuremberg 1947 answers the camp experiments. Belmont 1979 answers Tuskegee.

- Tuskegee, a US health service study, syphilis left untreated in Black men for forty years, penicillin withheld from 1947
- the press ended it in 1972, Congress answered with the National Research Act, Belmont followed

- Helsinki 1964 sits between, medicine writing its own rules, no single scandal behind it

First meeting, Brandt, the historian who anatomised Tuskegee:

“Only the press accounts of 1972 finally punctured the continued rationalizations of the USPHS and brought the study to an end.” (Brandt, 1978, p. 8)

Online research has its own scars.

- the T3 release fell to re-identification within days
- the contagion experiment, 689,003 news feeds tuned, nobody aware
- the journal that published it attached an Expression of Concern within weeks ([“Editorial Expression of Concern,” 2014](#))

The rules take decades to write. Online platforms have not existed that long.

- Nuremberg to Helsinki seventeen years, Helsinki to Belmont fifteen more, Facebook is from 2004
- internet research ethics is still versioning, the internet researchers’ association is on the third edition of its guidelines since 2002 ([franzke et al., 2020, p. 4](#))

The Nuremberg Code: “The voluntary consent of the human subject...”

- the first sentence of the first code, the camps are why it comes first

The National Statement: “On rare occasions...”

- the regulator keeps its origin story in its own preamble, the lineage runs unbroken to the document this unit sits under

Facebook, after the mood experiment: “Although this subject matter...”

- the modern echo, harm then rules, 2014 not 1947
- the rules it wrote were internal review, not law, a formal complaint to the American regulator went nowhere ([EPIC, 2014](#))

Cambridge Analytica, and Meta now

THE COST LANDS ON PEOPLE WHO NEVER AGREED

- Cambridge Analytica took data from 87 million people.
- Meta took every Australian public post since 2007 to train its AI models.
- Neither time did the people involved agree to that use.

Hu:

The FTC action, however, has been criticized as failing to adequately address the privacy and other harms emanating from Facebook’s release of approximately 87 million Facebook users’ data, which was exploited without user authorization.

([Hu, 2020, p. 1](#))

Ms Claybaugh, for Meta:

Correct. We use data that people have expressly made public

([Adopting Artificial Intelligence \(AI\), 2024, p. 2](#))

A Facebook user:

...With stuff like the recent Facebook scandal, it’s like you don’t realise how open your data is. I feel like a lot of companies probably do have my data now and I’ve just kind of got to the point where I’ve accepted, the basic data, I don’t care about sharing that with third parties anymore because I know most of them probably have it by this point.

([Hinds et al., 2020, p. 7](#))

Hu, on what the American regulator did:

The settlement, announced on 24 July 2019, included a record-setting \$5 billion

Cambridge Analytica took data from 87 million people.

- first meeting, Cambridge Analytica, a British political consulting firm, the harvested data fed voter profiling
- Kogan, the psychologist behind the quiz, framed it as academic research, the quiz reached the takers' friends too ([Hu, 2020, p. 2](#))
- the method sat in the open literature five years earlier, with its own consent warning ([Kosinski et al., 2013, p. 4](#))
- this cohort was about ten when it broke, history to them, not memory

Meta took every Australian public post since 2007 to train its AI models.

- on the record before the Australian Senate, 2024

The committee's count:

"there are millions of Australians who use Facebook and Instagram who have not consented to using their photos, their videos or the record of their lives and families to train an AI model." ([Adopting Artificial Intelligence \(AI\), 2024, p. 12](#))

Neither time did the people involved agree to that use.

- an externality, the cost lands on people outside the transaction
- AOL released users' search histories in 2006 believing them anonymised, people were re-identified, then T3, then the quiz, the same lesson unlearned ([Zimmer, 2010, p. 11](#))

Hu: "The FTC action, however..."

- first meeting, Hu, an American legal scholar on the case
- the FTC, America's consumer regulator, acted on the quiz
- the criticism, even a record fine under-addressed the harm

Ms Claybaugh, for Meta: "Correct. We use data..."

- first meeting, Melinda Claybaugh, Meta's privacy policy director, answering the Senate committee
- one word concedes the scrape, the qualifier does the work, expressly made public

A Facebook user: "...With stuff like the recent Facebook scandal..."

- first meeting, an interviewee from a 2020 study of users after the scandal ([Hinds et al., 2020](#))
- resignation, not consent, acceptance after the fact

Hu, on what the American regulator did: "The settlement..."

- five billion dollars says the harm was real, the order forces the consent the quiz never asked

The NHMRC against the world

EVERY JURISDICTION WRITES IT DIFFERENTLY

- Australian research ethics is anchored in statute.
- Two documents govern Australian researchers. The National Statement covers ethics. The Australian Code covers responsible conduct.
- American law leaves much of the public internet outside ethics review.
- The same study can require ethics approval in Australia and not in America.

The National Statement:

The National Health and Medical Research Council Act 1992 (Cth) (NHMRC Act) establishes NHMRC as a statutory body and sets out its functions, powers and obligations. Section 10(1) of the NHMRC Act requires the Chief Executive Officer (CEO) to issue human research guidelines... All the guidelines in this National Statement that are applicable to the conduct of research involving humans are issued by NHMRC in fulfilment of this statutory obligation.

([NHMRC et al., 2025, p. 7](#))

The NTIA:

Therefore, the secondary use of only non-identifiable data in research, for example, would generally not be subject to the Common Rule's requirements, even for research that is federally supported or conducted.

([NTIA, 2024, p. 5](#))

Australian research ethics is anchored in statute.

- the Statement exists because an Act requires it, section 10(1)
- funding is the enforcement lever

The Code's own teeth:

"Compliance with this Code is a requirement for the receipt of funding by NHMRC and ARC." ([NHMRC et al., 2018, p. 3](#))

Two documents govern Australian researchers.

- the Statement covers what may be done to people, the Code covers how researchers behave, integrity, authorship, data management
- both grow from the same Act, both enforced through funding, breach and the money stops
- an enforced code, not a professional courtesy

American law leaves much of the public internet outside ethics review.

- whole activities deemed not research at all, journalism, oral history, biography ([franzke et al., 2020, p. 14](#))
- common law reads online data as public, review negated

First meeting, Waldek and colleagues, Australian researchers on the ethics of hacked data, naming the split:

"This is very different to the USA and Canada where online data is frequently characterised under Common Law as public negating requirements for ethical review." ([Waldek et al., 2025, p. 4](#))

The same study can require ethics approval in Australia and not in America.

- same dataset, same method, different obligations
- the Australian system protects the person observed, the American system protects the freedom to inquire

The National Statement: “The National Health and Medical Research Council Act 1992...”

- statute on the screen, dry and load-bearing

The NTIA: “Therefore, the secondary use...”

- first meeting, the NTIA, a US Commerce Department agency
- the American hole stated in the American government’s own words

Four honest accounts

FOUR SYSTEMS ON COLLECTING WITHOUT CONSENT

Ethics review, the National Statement:

Importantly, the opt-out approach is unlikely to constitute consent when applying Commonwealth privacy legislation to the handling of sensitive information, including health information. Therefore, where it is impracticable to obtain an individual’s explicit consent to the use of their information... researchers must comply with the Guidelines under Section 95 of the Privacy Act 1988.

([NHMRC et al., 2025, p. 22](#))

Privacy law, the OAIC:

Just because data is publicly available or otherwise accessible does not mean it can legally be used to train or fine-tune generative AI models or systems

([OAIC, 2024, p. 2](#))

Research integrity, the Australian Code:

This Code does not incorporate the laws, regulations and guidelines and other codes of practice that apply to the conduct of research. Those responsible for the conduct of research are expected to be aware of and comply with the applicable laws and codes.

([NHMRC et al., 2018, p. 3](#))

Platform terms, Kogan’s defence:

Kogan contended that he conformed to Facebook’s guidelines at the time

([Hu, 2020, p. 2](#))

FOUR SYSTEMS ON COLLECTING WITHOUT CONSENT

- one question, collecting people’s data without their consent, four named rule systems speak to it, each protecting a different party
- the participant, the record, the citizen, the platform
- each speaks in its own register to its own cases, three lean no or defer, the platform’s register can say yes, and harm runs straight through a yes
- the OKCupid release, 2016, a Danish graduate student published a mass of dating-site profiles, the data already public his defence
- there the systems’ reactions landed for real and uncoordinated, the university distanced itself, copyright took the data down, the journal heard years later ([Waldek et al., 2025, p. 7](#))
- the defensible move, name the account followed and say why, the assessments mark exactly this

Ethics review, the National Statement: “Importantly, the opt-out approach...”

- protects the participant, opt-out consent fails the Privacy Act by itself
- the quiz takers’ friends never agreed at all, this account says no

Research integrity, the Australian Code: “This Code does not incorporate...”

- protects the record, the Code fences its own scope, the law still applies beside it
- an honest deferral, integrity does not decide the dataset question

Privacy law, the OAIC: “Just because data is publicly available...”

- first meeting, the OAIC, the Office of the Australian Information Commissioner, Australia's privacy regulator
- protects the citizen, public and legal-to-use are different questions, this account says no
- the verdict that reaches Meta's scrape, availability grants nothing

Platform terms, Kogan's defence: "Kogan contended that he conformed..."

- protects the platform, the terms granted what the quiz took
- the only yes among the four, and the harm ran straight through it

Who bears it, and who benefits?

RECAP B

- The rules were earned through harm.
- The same harm is running again today.
- One dataset can draw four honest verdicts.

The rules were earned through harm.

- the camps, Tuskegee, the codes, nothing in the sequence hypothetical

The same harm is running again today.

- the quiz then, the scrape now, the cost outside the transaction both times

One dataset can draw four honest verdicts.

- the OKCupid release drew all four, naming the account followed is the defensible move

Who bears it, and who benefits?

- the subjects bore it in every case named, the bearer never sat at the table

Public is not permission

REACHABLE IS A FACT ABOUT THE NETWORK

- Posting implies an audience. Researchers are rarely in it.
- Consent means agreeing to a specific use.
- An open account is not that agreement. Neither are unread terms.
- The AoIR guidelines are the internet research field's own ethics code, now in a third edition.

boyd and Crawford:

Just because content is publicly accessible does not mean that it was meant to be consumed by just anyone.

(Boyd & Crawford, 2012, p. 12)

Nissenbaum:

The notion that when individuals venture out in public, a street, a square, a park, a market, a football game, no norms are in operation, that "anything goes," is pure fiction.

(Nissenbaum, 2004, p. 22)

The National Statement:

Unless a waiver of the requirement for consent is obtained, any research access to or use of publicly available data or information must be in accordance with the consent obtained from the person to whom the data or information relates.

(NHMRC et al., 2025, p. 39)

The AoIR guidelines:

User-generated content is generally published in informal spaces that users often perceive as private but may strictly speaking be publicly accessible. In any case, researchers are rarely the intended audience of user-generated content.

(franzke et al., 2020, p. 70)

Posting implies an audience. Researchers are rarely in it.

- the cohort already practises audience design, the finsta, the close-friends list
- being in public and being public differ

boyd and Crawford drawing that line:

“Big Data researchers rarely acknowledge that there is a considerable difference between being in public (i.e. sitting in a park) and being public (i.e. actively courting attention).” (Boyd & Crawford, 2012, p. 13)

Consent means agreeing to a specific use.

- agreement travels with the use, not the account
- a hashtag invites a debate, not a dataset (Townsend & Wallace, 2016, p. 12)

First meeting, Nissenbaum, the philosopher of contextual integrity:

“personal information revealed in a particular context is always tagged with that context and never”up for grabs”” (Nissenbaum, 2004, p. 26)

An open account is not that agreement. Neither are unread terms.

- the contagion experiment claimed the data-use policy as informed consent (Kramer et al., 2014, p. 2)
- nobody reads the terms, measurably, just long enough to scroll to accept (Waldek et al., 2025, p. 5)

The AoIR guidelines are the internet research field’s own ethics code.

- the Association of Internet Researchers, the field regulating itself, three editions since 2002
- the enforcement ladder, AoIR guidance is voluntary, the Code and Statement enforce through funding, Europe’s privacy law the GDPR fines
- each rung binds differently, name the rung when naming a rule

boyd and Crawford: “Just because content is publicly accessible...”

- the field’s plainest sentence, access is not permission

Nissenbaum: “The notion that when individuals venture out in public...”

- pure fiction is her verdict on anything-goes in public space

The National Statement: “Unless a waiver...”

- the regulator’s version, use follows consent, waivers exist but must be argued

The AoIR guidelines: “User-generated content is generally published...”

- the audience rule, researchers are rarely the intended audience

Four readings of public

NO TWO OF THEM AGREE

The guiding principle for researchers is that, although data or information may be publicly available, this does not automatically mean that the individuals with whom this data or information is associated have necessarily granted permission for its use in research.

(NHMRC et al., 2025, p. 39)

Ms Claybaugh, for Meta:

That means when you go on Facebook or you go on Instagram and you make a post, you select the audience for that post—that statement, that photo, whatever it is you're posting online. If you choose to make that post, the text or the image, public, that is publicly sharing that information.

([Adopting Artificial Intelligence \(AI\), 2024, p. 2](#))

we believe that social media content, being publicly available and created voluntarily by users before the study, is free from ethical restraints

([Al-Zaman et al., 2024, p. 3](#))

A Facebook user:

We all are very exposed aren't we, in so many ways. Everything we do is exposed and we sell it to ourselves because most of the time they say, 'Oh if you like us and comment here and give us your Facebook, you will get a free cappuccino from Starbucks'... so everyone's going to give you all the data you want... Most people don't care about their privacy... and they don't even read the terms and conditions, they just click 'accept.'

([Hinds et al., 2020, p. 7](#))

NO TWO OF THEM AGREE

- one word, public, four working definitions in the wild right now

The National Statement: “The guiding principle for researchers...”

- the baseline reading, availability grants nothing by itself

al-Zaman and colleagues: “we believe that social media content...”

- first meeting, al-Zaman and colleagues, media researchers publishing in 2024
- a live position in print, not a straw man, the OKCupid student held the same reading ([Waldek et al., 2025, p. 7](#))

Ms Claybaugh, for Meta: “That means when you go on Facebook...”

- publicness as audience selection, selection read as consent to a different use
- the regulator reads the same silence the other way

A Facebook user: “We all are very exposed aren't we...”

- the cappuccino trade, the price visible and tiny, the cost invisible and cumulative
- resignation reads as consent only from outside, the shrug is the strongest evidence the terms are not agreement

Reading a policy for its hole

WHO IT LEAVES UNPROTECTED

- Reading a policy means looking for what it fails to cover.
- The OAIC guidance is addressed to developers, not to the people scraped.
- It names no remedy for them.

The OAIC, 21 October 2024:

a failure of a website to implement measures to prevent data scraping should not be taken as implied consent

([OAIC, 2024](#))

Ms Claybaugh, for Meta, 11 September 2024:

We do provide an opt out so that people can say, 'I no longer want my public posts and images to be used to train the models,' so we do provide that opt out.

([Adopting Artificial Intelligence \(AI\), 2024, p. 8](#))

Senator Shoebridge, to Meta:

she would never have contemplated that Meta was going to scrape those photos ... and yet you chose to just sweep all that information into your AI model without asking her consent. Can't you see the ethical problem there?

(Adopting Artificial Intelligence (AI), 2024, p. 4)

Reading a policy means looking for what it fails to cover.

- the method, quote it, say what it covers, say who falls outside
- the pair on screen sits 40 days apart, one country, one dataset, no overlap in protection

The OAIC guidance is addressed to developers, not to the people scraped.

- the OAIC again, this time writing guidance for AI developers
- the address line decides who a document protects, the scraped appear only as the data

The document's own address line:

"This guidance is intended for developers of generative artificial intelligence (AI) models or systems who are subject to the Privacy Act" *(OAIC, 2024, p. 6)*

It names no remedy for them.

- nobody scraped can opt out after the fact, the scrape already happened
- guidance is not a clause, no penalty, no duty on anyone

The OAIC, 21 October 2024: "a failure of a website..."

- the regulator reads silence as no consent

Ms Claybaugh, for Meta, 11 September 2024: "We do provide an opt out..."

- the platform reads an unticked box as a green light
- the question was about the taking, the answer is about output filters

Senator Shoebridge, to Meta: "she would never have contemplated..."

- first meeting, Senator Shoebridge, Greens, leading the committee's questioning
- his question is about a mother who posted family photos, Meta swept them into its AI training
- she is the person in the hole, absent from every protection on the slide

De-identified is a claim

A QUOTE IS A SEARCH KEY

- De-identified describes a process. It is not a permanent state.
- A verbatim quote leads a search engine straight back to the person.
- The Berkeley Protocol is the United Nations manual for open-source investigation. Its standards anchor this movement.

Townsend and Wallace:

Of particular concern is the republishing of quotes that have been taken from social media platforms and republished verbatim, as these can lead us, via search engines, straight back to their original location, often then exposing the identity and profile of the social media user they originate from.

(Townsend & Wallace, 2016, p. 7)

Taylor and colleagues:

Google immediately identified the social media influencer in all de-identified trials, despite the anonymization techniques we had used. This indicated that traditional anonymization strategies for visual print data were, in fact, not effective in an online context.

(Taylor et al., 2023, p. 4)

De-identified describes a process. It is not a permanent state.

- the regulator refuses the category, a process can be undone ([NHMRC et al., 2025, p. 36](#))
- three fields identify most Americans, birth date, gender, postcode ([Zimmer, 2010, p. 7](#))

Zimmer on the T3 fall:

“Despite these steps, and claims by the T3 researchers that”all identifying information was deleted or encoded”, the identity of the source of the dataset was quickly discovered.” ([Zimmer, 2010, p. 1](#))

A verbatim quote leads a search engine straight back to the person.

- the mechanism is a string search, nothing clever ([franzke et al., 2020, p. 11](#))
- paraphrase is the working defence, ethnographers adopted it before search engines demanded it ([Boellstorff et al., 2012, p. 12](#))

The Berkeley Protocol is the United Nations manual for open-source investigation.

- the UN’s standard for court-grade online investigation, 2022
- one manual, two pages, two positions, the definition promises impossibility, the mosaic effect concedes it

Townsend and Wallace: “Of particular concern is the republishing...”

- first meeting, Townsend and Wallace, authors of Aberdeen’s social media research ethics guide
- a republished quote is a search key leading straight back to the person

Taylor and colleagues: “Google immediately identified...”

- first meeting, Taylor and colleagues, an empirical test in 2023
- every de-identified trial fell to one search

The Berkeley Protocol, page 90: “Anonymization: the process...”

- the promise

The Berkeley Protocol, page 25: “Investigators should also be aware...”

- the concession, two pages from the promise

The ethics of hacked data

SANCTIONED IS NOT FREE

- Research on hacked data exists. Sometimes it is approved.
- The conditions of approval show how high the cost is.

Waldek and colleagues:

Additional safeguarding around the process of engagement with the data were also put in place, including time limits for viewing the data, mental health

- This unit's path does not go there.

first aid assessment processes and trauma informed practice techniques.

([Waldek et al., 2025, p. 14](#))

The AoIR guidelines:

when is it allowable (if ever) to use data that would otherwise be prohibited ethically and/or legally because of privacy protections, etc., but has been made public because of an accidental breach and/or intentional hack

([franzke et al., 2020, p. 20](#))

Research on hacked data exists. Sometimes it is approved.

- the Waldek paper is my own team's work, approval on the record, HRE23-177 ([Waldek et al., 2025, p. 18](#))
- Australia mandates committee approval here, the Privacy Act defines the sensitivity

The conditions of approval show how high the cost is.

- the safeguards were personal, not procedural
- consent itself was weighed and set aside, asking judged to re-harm the violated

Waldek's own weighing:

"in our case we believe seeking consent would arguably do more harm." ([Waldek et al., 2025, p. 13](#))

This unit's path does not go there.

- nothing in the assessments touches hacked material, the case shows where the edge is
- using leaked data and exploiting leaked credentials differ, the second prohibited outright ([OHCHR & Human Rights Center, 2022, p. 41](#))

Waldek and colleagues: "Additional safeguarding..."

- each safeguard on the list is a cost somebody paid

The AoIR guidelines: "when is it allowable (if ever)..."

- the field's open question against the concrete price beside it
- our paper answers that question directly for one case

Four positions on hacked data

JUST BECAUSE WE HAVE IT DOES NOT MEAN WE CAN USE IT

The National Statement:

This includes avoiding the use or disclosure of information that was obtained unethically or illegally.

([NHMRC et al., 2025, p. 40](#))

Poor and Davidson:

we want this data but we don't need it

(Poor and Davidson 2016, quoted in ([Waldek et al., 2025, p. 11](#)))

Post:

to use [hacked] data without the consent of those who were violated is to violate the violated anew

(Post 1991, quoted in ([Waldek et al., 2025, p. 13](#)))

An open-source journalist:

The importance of the investigative goal is often greater than the legality of the means.

([Van Der Woude et al., 2025, p. 10](#))

JUST BECAUSE WE HAVE IT DOES NOT MEAN WE CAN USE IT

- four positions, refusal on principle, refusal on sufficiency, the regulator's narrow door, the working journalist

The National Statement: "This includes avoiding the use or disclosure..."

- the regulator's default, an exception must argue against it

The narrow door in the same document:

"Researchers should establish that these risks are justified by the benefits of the research, which may include the exposure itself; for example, where the illegal activity bears on the discharge of a public responsibility or the fitness to hold public office." ([NHMRC et al., 2025, p. 102](#))

Post: "to use [hacked] data without the consent..."

- first meeting, Post, writing in 1991, before the internet cases
- refusal on principle, consent missing twice, the hack and the research

Poor and Davidson: "we want this data but we don't need it"

- first meeting, Poor and Davidson, refusing hacked Patreon data in 2016
- refusal on sufficiency, wanted but not needed fails the weighing on its face

An open-source journalist: "The importance of the investigative goal..."

- first meeting, an anonymised working journalist from a 2025 interview study
- the story over the legality, the weighing continuous, no umpire sits above the four

Who bears it, and who benefits?

RECAP C

- Public is not permission.
- A verbatim quote can undo a de-identification claim.
- Hacked-data research is sometimes approved. Sanctioned is not free.

Public is not permission.

- four readings of one word could not agree, the strongest evidence was the user's shrug

A verbatim quote can undo a de-identification claim.

- the search engine does the re-identifying, paraphrase is the working defence

Hacked-data research is sometimes approved.

- the Waldek conditions show the price of a yes

Who bears it, and who benefits?

- the re-identified bore it each time, the assessments ask the same question back

Collect exactly what you need

COLLECTION IS WHERE OBLIGATION BEGINS

- Collect exactly what the question needs and nothing more.

Cegłowski:

- Once data is collected, the obligation to guard it begins.
- Material handed to an AI service is a disclosure that cannot be undone.

In particular, I'd like to draw a parallel between what we're doing and nuclear energy, another technology whose beneficial uses we could never quite untangle from the harmful ones.

([Cegłowski, 2015, p. 1](#))

Cegłowski:

A singular problem of nuclear power is that it generated deadly waste whose lifespan was far longer than the institutions we could build to guard it.

([Cegłowski, 2015, p. 1](#))

The AoIR guidelines:

Once an AI system has left the hands of the original researchers, they may not have any control over how their models are used by others. The same is true for the generated research data: once it has been freely published, it will be difficult to contain its further uses.

([franzke et al., 2020, p. 46](#))

Collect exactly what the question needs and nothing more.

- deciding what to collect is where the ethics starts ([NHMRC et al., 2025, p. 37](#))
- the Protocol's three-part test, justified, necessary, proportional ([OHCHR & Human Rights Center, 2022, p. 26](#))

Once data is collected, the obligation to guard it begins.

- the Moscow diary again, collected in one world, dangerous in the next
- holding data is a liability, not an asset

Over-collection in the Protocol's words:

"Over-collection, a particular concern when using automated collection processes, may create or exacerbate security vulnerabilities, in particular if it leads to investigators being unaware of the types of information within their possession." ([OHCHR & Human Rights Center, 2022, p. 27](#))

Material handed to an AI service is a disclosure that cannot be undone.

- the clipboard is the disclosure event, deletion cannot be compelled afterwards
- whether a pasted transcript trains a model is the provider's setting, not the researcher's choice ([NHMRC, 2025, p. 6](#))

Cegłowski: "In particular, I'd like to draw a parallel..."

- the parallel is his argument, not a decoration, benefits never untangled from harms

Cegłowski: "A singular problem of nuclear power..."

- held data has the same lifespan problem, the waste outlives its guards

The AoIR guidelines: "Once an AI system has left the hands..."

- control ends at publication, obligation does not

Our obligations around research data

AS OPEN AS POSSIBLE AS CLOSED AS NECESSARY

Rather, the ORD pilot follows the principle “as open as possible, as closed as necessary” and focuses on encouraging sound data management as an essential part of research best practice.

([European Commission, 2016, p. 4](#))

The NHMRC’s data guide:

for areas such as gene therapy, research data must be retained permanently (e.g. data in the form of patient records).

([NHMRC et al., 2019, p. 6](#))

If you have to store it, don’t keep it!

([Cegłowski, 2015, p. 9](#))

The National AI Plan:

AI models are only as good as the data they are trained on... large, unstructured datasets could be made accessible for AI system training.

([DISR, 2025, p. 15](#))

AS OPEN AS POSSIBLE AS CLOSED AS NECESSARY

- four answers to one question, what is owed to the data we hold

The European Commission: “Rather, the ORD pilot...”

- first meeting, the European Commission’s funding rules for open research data
- open to accelerate research, closed to protect people, my own practice sits closer to the open side
- trouble arrives when the data is other people’s posts, those people alive and findable

Cegłowski: “If you have to store it, don’t keep it!”

- deletion argued as the ethical maximum, not a failure ([Markham et al., 2018, p. 6](#))

The NHMRC’s data guide: “for areas such as gene therapy...”

- first meeting, the NHMRC’s data management guide, the Statement’s sibling for records
- retention is real law for some classes, the record outlives the project by design, integrity is the party protected

The National AI Plan: “AI models are only as good...”

- first meeting, the government’s 2025 AI plan, direction rather than statute
- current policy wants the data flowing the other way, the disagreement is genuine and current

Every investigation serves someone

THE JOB COMES WITHOUT A RULEBOOK

- Courts, media outlets and agencies commission this work.
- Professional investigators often work without an ethics committee.
- Whoever does the work ends up writing the rules.

Van der Woude and colleagues:

Open source investigators’ narratives reveal that their work relies on informal peer control and their own personal adherence to the “do not harm” and “public interest” principles rather than a professionally and widely agreed code of ethics.

([Van Der Woude et al., 2025, p. 14](#))

The Berkeley Protocol:

Open source investigators must be accountable for their actions, which can often be ensured through clear documentation, record-keeping and oversight.

([OHCHR & Human Rights Center, 2022, p. 24](#))

The National Statement:

The researcher is responsible and accountable to their institution, any sponsors or funders of the research, participants and, in some research, to regulators or other entities who have a formal role in the oversight of the research.

([NHMRC et al., 2025, p. 117](#))

An open-source journalist:

Courts, media outlets and agencies commission this work.

- the commissioning bodies are named in the sources ([OHCHR & Human Rights Center, 2022, p. 8](#))
- the client shapes the question, Meta's scrape was a commercial decision, no committee in sight ([Adopting Artificial Intelligence \(AI\), 2024, p. 4](#))

Professional investigators often work without an ethics committee.

- accountability shifts to documentation and cross-examination, methods must survive a courtroom
- the newsroom convenes no standing committee, colleagues and conscience, continuously

Whoever does the work ends up writing the rules.

- few of the jobs this degree leads to come with a committee
- communications, policy, consulting, each will ask for a look at an online space eventually

Van der Woude and colleagues: "Open source investigators' narratives..."

- first meeting, van der Woude and colleagues, a 2025 interview study of open-source journalists
- informal peer control standing in for a professional code that does not exist

The Berkeley Protocol: "Open source investigators must be accountable..."

- the Protocol's answer, documentation, record-keeping, oversight, testimony at trial

The National Statement: "The researcher is responsible and accountable..."

- the academic's scaffolding listed in one sentence, institution, funder, participant, regulator

An open-source journalist: "We don't have any guidelines..."

- the practitioner wants rules that do not exist, somebody will write them

The rules for Assessment 1

ANALYSE THE PLATFORM, NOT THE PEOPLE

- Assessment 1 analyses a platform, not the people on it.
- The brief asks for the least data that answers the question.
- These rules are one defensible judgement among several possible ones.

Assessment 1 analyses a platform, not the people on it.

- the object is the platform's design, affordances, defaults, interaction patterns
- the people stay unstudied, platform choice lands in tutorial

The brief asks for the least data that answers the question.

- collect exactly what you need, applied to coursework
- the affordance table and the diagram collect nobody, a screenshot is the exception that argues itself, anonymised before it enters the report

These rules are one defensible judgement among several possible ones.

- the four honest accounts would each write this brief differently
- disagreement with reasons is welcome, arguing against the rules well is the skill being marked

Arguing inside an open question

THE LAW IS MOVING WHILE THE ESSAYS ARE WRITTEN

- The frameworks genuinely disagree.
- Australian law is under review right now.
- The skill is arguing a defensible position inside an open question.

The AoIR guidelines:

*the issues raised by Internet research are ethical problems precisely because they evoke more than one ethically defensible response to a specific dilemma or problem. **Ambiguity, uncertainty, and disagreement are inevitable.***

([franzke et al., 2020, p. 7](#))

The National AI Plan:

Australia has strong protections in place to address many risks, but the technology is fast-moving and regulation must keep pace. That's why the government continues to assess the suitability of existing laws in the context of AI.

([DISR, 2025, p. 29](#))

The frameworks genuinely disagree.

- more than one defensible answer is the normal condition, the field's guides conflict with each other ([Townsend & Wallace, 2016, p. 4](#))
- harm-based and dignity-based privacy read one dataset differently

Zimmer naming the split behind slide one's argument:

"Such a position supposes that so long as the data can be protected from attack by hackers or others wishing to"do" something harmful once gaining access, the privacy of the subjects can be maintained. Such a position ignores the broader dignity-based theory of privacy." ([Zimmer, 2010, p. 9](#))

Australian law is under review right now.

- the AI plan promises iteration, not settlement ([DISR, 2025, p. 8](#))
- the essays fall due while it moves

The skill is arguing a defensible position inside an open question.

- three moves, a position, its account named, its cost-bearer named
- the law will move again, the method survives the movement ([franzke et al., 2020, p. 15](#))

The AoIR guidelines: "the issues raised by Internet research..."

- the bold sentence, ambiguity is inevitable, stated not regretted

The National AI Plan: "Australia has strong protections..."

- the government saying the question is open, in its own words

References

Adopting Artificial Intelligence (AI) (September 11, 2024).

[https://parlinfo.aph.gov.au/parlInfo/download/committees/commsen/28391/toc_pdf/Adopting%20Artificial%20Intelligence%20\(AI\)%20Select%20Committee_2024_09_11.pdf;fileType=application%2Fpdf](https://parlinfo.aph.gov.au/parlInfo/download/committees/commsen/28391/toc_pdf/Adopting%20Artificial%20Intelligence%20(AI)%20Select%20Committee_2024_09_11.pdf;fileType=application%2Fpdf)

Al-Zaman, Md. S., Khemka, A., Zhang, A., & Rockwell, G. (2024). The defining characteristics of ethics papers on social media research: A systematic review of the literature. *Journal of Academic Ethics*, 22(1), 163–189. <https://doi.org/10.1007/s10805-023-09491-7>

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). Ethics. In *Ethnography and Virtual Worlds* (pp. 129–150). Princeton University Press. <https://doi.org/10.2307/j.cttq9s20.12>

Boyd, D., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>

Brandt, A. M. (1978). Racism and Research: The Case of the Tuskegee Syphilis Study. *The Hastings Center Report*, 8(6), 21. <https://doi.org/10.2307/3561468>

Ceglowski, M. (2015). *Haunted By Data*. Strata+Hadoop World Conference, New York. https://idlewords.com/talks/haunted_by_data.htm

DISR. (2025, December 2). *National AI Plan*. <https://www.industry.gov.au/sites/default/files/2025-12/national-ai-plan.pdf>

Editorial Expression of Concern: Experimental evidence of massivescale emotional contagion through social networks. (2014). *Proceedings of the National Academy of Sciences*, 111(29), 10779–10779. <https://doi.org/10.1073/pnas.1412469111>

EPIC. (2014, July). *In re: Facebook (Psychological Study)*. EPIC - Electronic Privacy Information Center. <https://epic.org/documents/in-re-facebook-psychological-study/>

European Commission. (2016). *Guidelines on FAIR Data Management in Horizon 2020* (Version 3.0).

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

franzke, aline shakti, Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers. (2020). *Internet Research: Ethical Guidelines 3.0*.

<https://aoir.org/reports/ethics3.pdf>

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575–599. <https://doi.org/10.2307/3178066>

Hinds, J., Williams, E. J., & Joinson, A. N. (2020). “It wouldn’t happen to me”: Privacy concerns and perspectives following the cambridge analytica scandal. *International Journal of Human-Computer Studies*, 143, 102498. <https://doi.org/10.1016/j.ijhcs.2020.102498>

Hu, M. (2020). Cambridge analytica’s black box. *Big Data & Society*, 7(2), 2053951720938091. <https://doi.org/10.1177/2053951720938091>

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802–5805. <https://doi.org/10.1073/pnas.1218772110>

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790. <https://doi.org/10.1073/pnas.1320040111>

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4), 330–342. <https://doi.org/10.1016/j.socnet.2008.07.002>

Markham, A. N., Tiidenberg, K., & Herman, A. (2018). Ethics as Methods: Doing Ethics in the Era of Big Data Research—Introduction. *Social Media + Society*, 4(3), 2056305118784502. <https://doi.org/10.1177/2056305118784502>

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report*.

- NHMRC. (2025, December). *Guide for assessing research involving Artificial Intelligence, Machine Learning and Large Language Model Technology* (collectively “AI”).
<https://www.nhmrc.gov.au/about-us/resources/guide-assessing-research-involving-artificial-intelligence-machine-learning-and-large-language-model-technology-collectively-ai>
- NHMRC, Australian Research Council, & Universities Australia. (2018). *The Australian Code for the Responsible Conduct of Research*.
- NHMRC, Australian Research Council, & Universities Australia. (2019). *Management of Data and Information in Research: A guide supporting the Australian Code for the Responsible Conduct of Research*.
- NHMRC, Australian Research Council, & Universities Australia. (2025). *National Statement on Ethical Conduct in Human Research 2025*.
- Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, 79.
- NTIA. (2024, December 11). *Ethical guidelines for research using pervasive data*. Federal Register.
<https://www.federalregister.gov/documents/2024/12/11/2024-29064/ethical-guidelines-for-research-using-pervasive-data>
- OAIC. (2024, October 21). *Guidance on privacy and developing and training generative AI models*. OAIC. <https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/guidance-on-privacy-and-developing-and-training-generative-ai-models>
- OHCHR, & Human Rights Center, S. of L., Berkeley. (2022). *Berkeley Protocol on Digital Open Source Investigations: A Practical Guide on the Effective Use of Digital Open Source Information in Investigating Violations of International Criminal, Human Rights and Humanitarian Law*. OHCHR. <https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source>
- Schroepfer, M. (2014, October 2). *Research at Facebook*. Meta Newsroom.
<https://about.fb.com/news/2014/10/research-at-facebook/>
- Taylor, N., Valencia-García, L. D., VandenBroek, A., Stinnett, A., & Allen, A. (2023). Ethics and images in social media research. *First Monday*. <https://doi.org/10.5210/fm.v28i4.12680>
- The Nuremberg Code (1947). (1996). *BMJ*, 313(7070), 1448.1.
<https://doi.org/10.1136/bmj.313.7070.1448>
- Townsend, D. L., & Wallace, C. (2016). *Social Media Research: A Guide to Ethics*.
- Van Der Woude, M., Dodds, T., & Torres, G. (2025). The ethics of open source investigations: Navigating privacy challenges in a gray zone information landscape. *Journalism*, 26(10), 2184–2202. <https://doi.org/10.1177/14648849241274104>
- Waldek, L., Ballsun-Stanton, B., Iqbal, M., Kernot, D., & Smith, D. (2025). Ethical conundrums: Hacked data in the study of far-right violent extremism. *New Media & Society*, 14614448251399640. <https://doi.org/10.1177/14614448251399640>
- Zimmer, M. (2008, October 1). *On the “Anonymity” of the Facebook Dataset (Updated)*. MichaelZimmer.org. <https://michaelzimmer.org/2008/09/30/on-the-anonymity-of-the-facebook-dataset/>
- Zimmer, M. (2010). “But the data is already public”: On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313–325. <https://doi.org/10.1007/s10676-010-9227-5>

AI disclosure

- These Week 3 materials were built with Claude (Anthropic): about 2.6 million tokens of AI output, steered by about 121,000 words of human direction across INTS1302 to date.
- Your assessments ask you to state which AI tools you used and how. This slide is that statement, for ours.