# AI and Ethics

Richard Wheeler

Reykjavik University

# Let’s Talk Ethics and Processes.

In the real world, ethics often means being compliant with laws, regulations, and best practices. It is about avoiding lawsuits, as few companies have any interest in being *moral*. A *moral* company is called a *charity* and have their own rules and culture.

Some fields have special laws and guidelines for everything, including medicine, public health, energy, etc. Other fields are mostly unregulated, and ethics is defined by what you can get away with.

# Let’s Talk Ethics and Processes.

In university and research contexts, ethics are easier, as ethical rules and guidelines are usually codified in *Rules of Ethical Conduct* or something similar, and so you at least have a guide to follow. Universities usually have an ethics committee that quarterly reviews cases put forward by their governing and legal bodies.

Cases often are about:

- Approving research work that involved medicine or clinical trials
- Academic misconduct: falsifying results, faking publications, taking credit for another’s work, etc.
- Conflicts between staff, and inappropriate behaviour in the workplace
- Introduction of new guidelines against bias and equal opportunities for staff and students
- Review of disabled persons’ access, wrongful termination of staff, wrongful dismissal or failing of students…

# Let’s Talk Ethics and Processes.

Ethical review boards are usually tasked with giving feedback to senior administrators on whether behaviour was ethical or not, and whether there is likely to be legal exposure to consider.

It’s very rare for a university to review:

- The robustness or appropriateness of AI systems in use or AI research and methods
- Whether research conducted may have ethical considerations outside of the fields of medicine, biology, etc., or unless it is specifically identified as having ethical issues

Within a research project that has any ethical issues, it is best practice to establish an ethics review committee that reviews all research activities, especially where there might be contact or influence on external persons. The committee should be reviewing all such work, which can only proceed once the ethical review committee agrees that such work complies with the highest ethical standards.

# Let’s Explore Some Examples

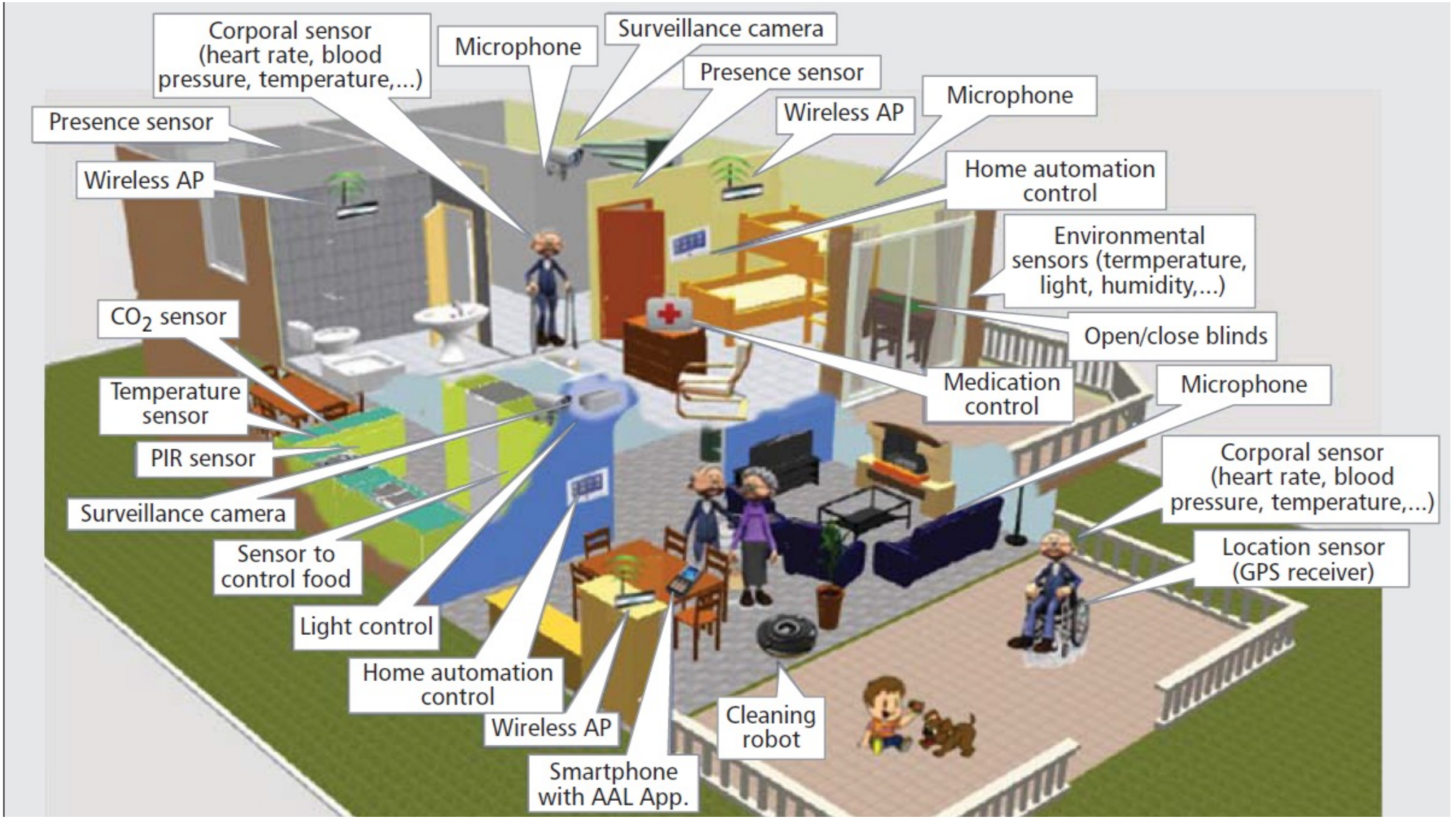
Let’s discuss some examples.  These are all from a single project: *SAAM – Ambient Intelligence for Assisted Living.* The project designed ambient monitoring systems for the elderly in Europe’s poorest countries (https://saam2020.eu/).

# Example: *SAAM*

*SAAM* was designed to monitor elderly persons living alone in some of Europe's poorest regions. For example, the system needed to be able to know if an elderly person fell and could not get up, or if their behaviour changed radically, or if they did not get out of bed in the morning, or if they were up all night. The system was tasked with learning an individual older person's habits and routines, and then issuing a warning of some kind in the event of trouble, for example, notifying a neighbour, relative, friend, or even the Bulgarian Red Cross or medical services.

# Example: *SAAM*

This field of helping seniors and others in their home through ICT is often called *ambient intelligence,* or *ambient assisted living (AAL).* AAL usually looks like this…


Corporal sensor (heart rate, blood pressure, temperature,...)
Microphone
Surveillance camera
Presence sensor
Wireless AP
Microphone
Presence sensor
Wireless AP
Home automation control
Environmental sensors (termperature, light, humidity,...)
$CO_2$ sensor
Open/close blinds
Temperature sensor
Medication control
Microphone
PIR sensor
Corporal sensor (heart rate, blood pressure, temperature,...)
Surveillance camera
Location sensor (GPS receiver)
Sensor to control food
Light control
Home automation control
Wireless AP
Cleaning robot
Smartphone with AAL App.

# But in Bulgaria


Каритас

# What can we do for her?

# Example: *SAAM*

We formulated SAAM to develop AAL technology for the very lowest economic grouping, with a target of €15/year per user.

For this reason, we cannot use special sensors and sensing arrays, robots, laptop computers or tablets, or special ICT. So, no cameras, no microphones, *no mobile phones, no WIFI routers,* no smart watches or bands, no sleep sensors, no IoT.

We need to support users independent of language, and provide support without any special hardware.

# Ideas for how to create such a system?

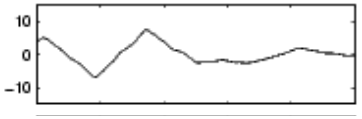
# Smart Meters

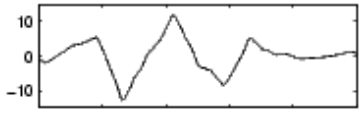
Smart meters for electricity are common in the EU, and are now standard, *especially* in poorer areas. They collect and send real-time information about energy use in the home.

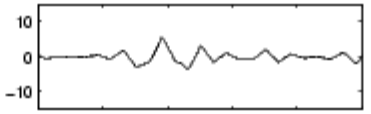
# Smart Meters







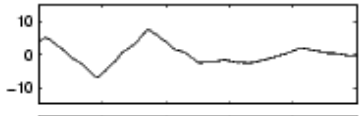




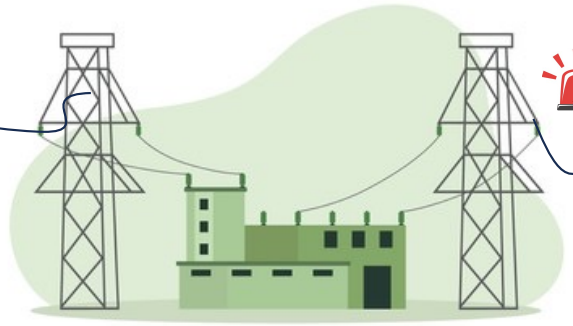

Unique electrical signature = activity profile

# Smart Meters

We know something is wrong because she never cooked this morning or made coffee.

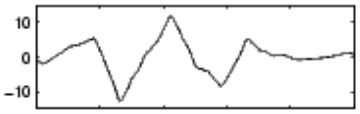


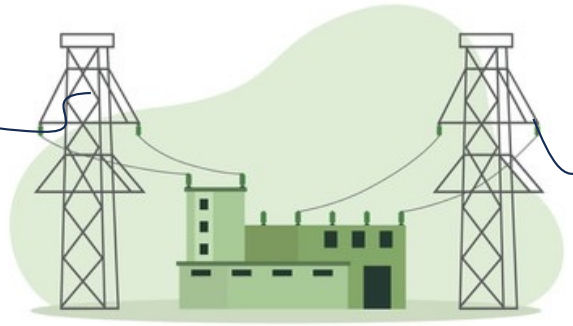




# Smart Meters

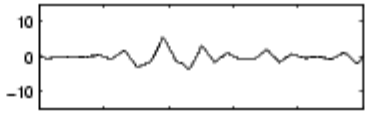
We know something is wrong because she was in the bathroom six times in the night, and never turned off the bedroom light after the last bathroom visit.



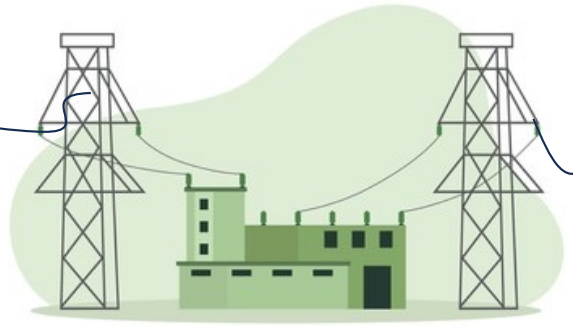



# Smart Meters

We know something is wrong because she has stopped cooking, and the TV has been left on all day and night.







# Let’s Explore Ethical Examples

I was the primary architect, but then did none of the technical work myself, only specified it to the consortium of universities, etc.

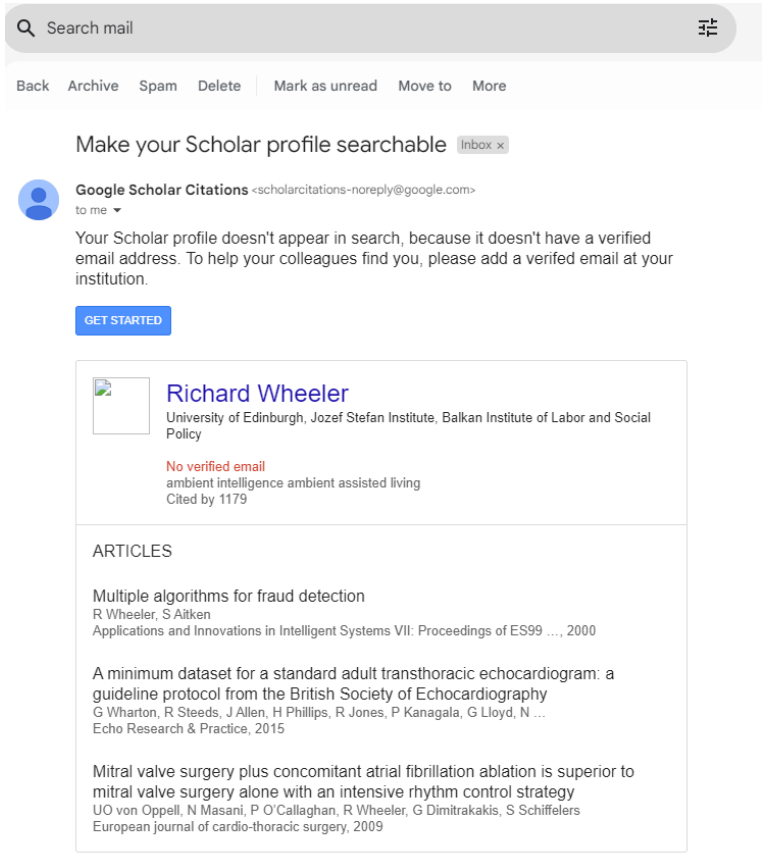
**First case, academic integrity**: many of the publications in the project listed me as author because I did the master design of nearly every AI and ML component, the architecture, etc.  But I did not contribute to these papers at all, some I did not even read. Is that ethical?

# Let’s Explore Some Examples

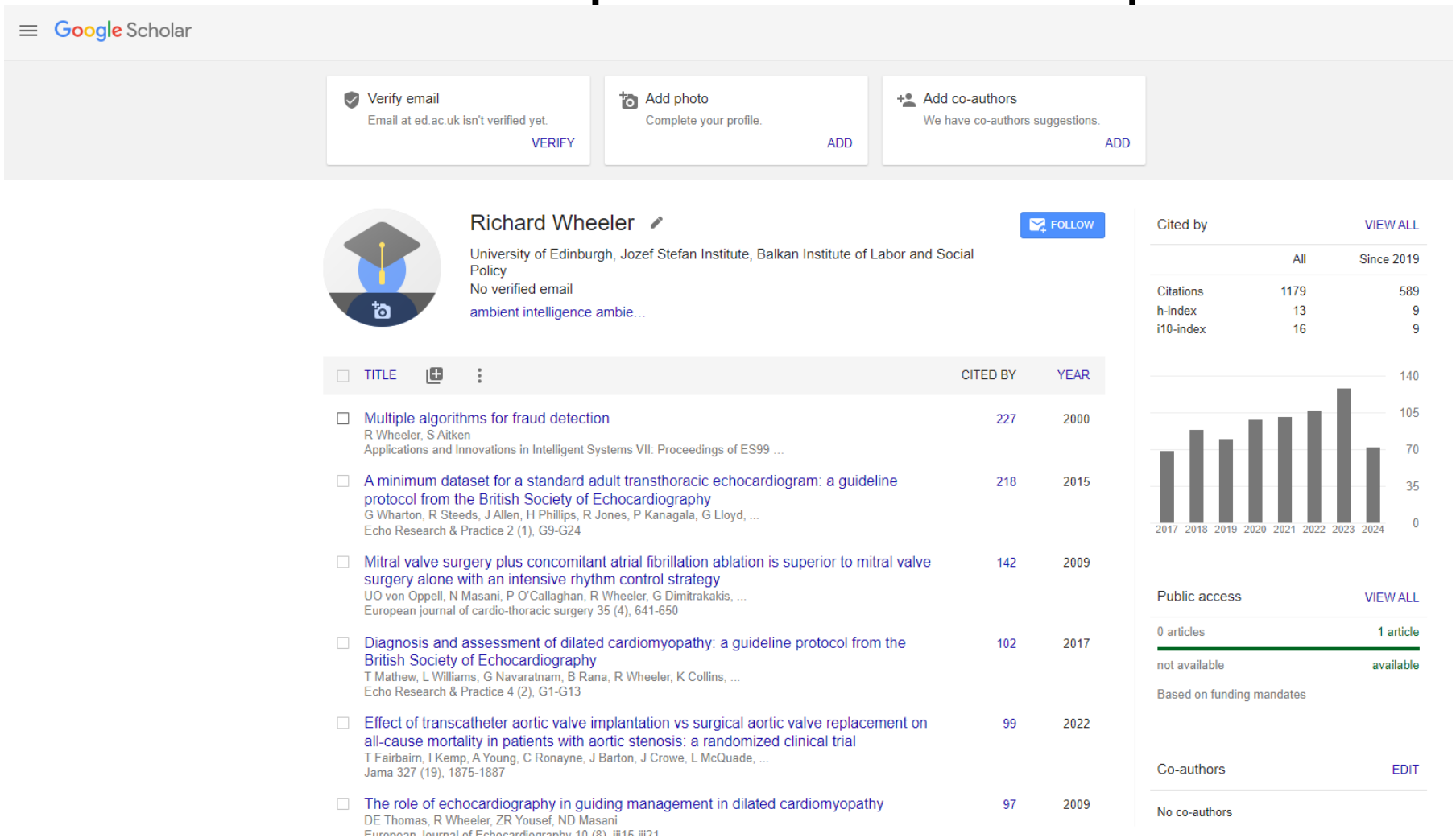
**Second case, academic integrity**: many of the publications in the project *did not* list me as author *although* I did the master design of nearly every AI and ML component, the architecture, etc. I got to see publications of my ideas, architectures, graphics, texts from the proposal, etc., published in other people’s names, with no credit to me. Is that ethical?

# Speaking of SAAM Publications and Attribution…

**Third case, academic system integrity**:



# Let’s Explore Some Examples



# Let's Explore Some Examples

| TITLE | CITED BY | YEAR |
|---|---|---|
| Multiple algorithms for fraud detection<br>R Wheeler, S Aitken<br>Applications and Innovations in Intelligent Systems VII: Proceedings of ES99 ... | 227 | 2000 |
| A minimum dataset for a standard adult transthoracic echocardiogram: a guideline protocol from the British Society of Echocardiography<br>G Wharton, R Steeds, J Allen, H Phillips, R Jones, P Kanagala, G Lloyd, ...<br>Echo Research & Practice 2 (1), G9-G24 | 218 | 2015 |
| Mitral valve surgery plus concomitant atrial fibrillation ablation is superior to mitral valve surgery alone with an intensive rhythm control strategy<br>UO von Oppell, N Masani, P O'Callaghan, R Wheeler, G Dimitrakakis, ...<br>European journal of cardio-thoracic surgery 35 (4), 641-650 | 142 | 2009 |
| Diagnosis and assessment of dilated cardiomyopathy: a guideline protocol from the British Society of Echocardiography<br>T Mathew, L Williams, G Navaratnam, B Rana, R Wheeler, K Collins, ...<br>Echo Research & Practice 4 (2), G1-G13 | 102 | 2017 |
| Effect of transcatheter aortic valve implantation vs surgical aortic valve replacement on all-cause mortality in patients with aortic stenosis: a randomized clinical trial<br>T Fairbairn, I Kemp, A Young, C Ronayne, J Barton, J Crowe, L McQuade, ...<br>Jama 327 (19), 1875-1887 | 99 | 2022 |
| The role of echocardiography in guiding management in dilated cardiomyopathy<br>DE Thomas, R Wheeler, ZR Yousef, ND Masani<br>European Journal of Echocardiography 10 (8), iii15-iii21 | 97 | 2009 |
| A minimum dataset for a standard transoesophageal echocardiogram: a guideline protocol from the British Society of Echocardiography<br>R Wheeler, R Steeds, B Rana, G Wharton, N Smith, J Allen, J Chambers, ...<br>Echo Research & Practice 2 (4), G29-G45 | 41 | 2015 |
| Guidelines for transoesophageal echocardiographic probe cleaning and disinfection from the British Society of Echocardiography<br>P Kanagala, C Bradley, P Hoffman, RP Steeds<br>European journal of echocardiography 12 (10), i17-i23 | 38 | 2011 |
| Stress echocardiography in coronary artery disease: a practical guideline from the British Society of Echocardiography<br>RP Steeds, R Wheeler, S Bhattacharyya, J Reiken, P Nihoyannopoulos, ...<br>Echo Research & Practice 6 (2), G17-G33 | 35 | 2019 |
| Charge-coupled devices for the ESA Euclid M-class mission<br>J Endicott, S Darby, S Bowring, D Burt, T Eaton, A Grey, I Swindells, ...<br>High Energy, Optical, and Infrared Detectors for Astronomy V 8453, 19-26 | 33 | 2012 |
| The role of echocardiography in the management of atrial fibrillation<br>R Wheeler, ND Masani<br>European Journal of Echocardiography 12 (10), i33-i38 | 26 | 2011 |
| The Euclid VIS CCD detector design, development, and programme status<br>AD Short, D Barry, M Berthe, N Boudin, O Boulade, R Cole, M Cropper, ...<br>High Energy, Optical, and Infrared Detectors for Astronomy VI 9154, 234-246 | 23 | 2014 |
| A systematic approach to echocardiography in hypertrophic cardiomyopathy: a guideline protocol from the British Society of Echocardiography<br>N Smith, R Steeds, N Masani, J Sandoval, G Wharton, J Allen, ...<br>Echo Research & Practice 2 (1), G1-G7 | 17 | 2015 |
| A safety checklist for transoesophageal echocardiography from the British Society of Echocardiography and the Association of Cardiothoracic Anaesthetists | 13 | 2015 |

**How many of you think:**

- Google got this wrong because our names are similar – it's a mix-up?
- Google *engineered* this to be wrong so I would do free work for them and engage with their platform?
- Google *engineered* this to be wrong so they could sell my data to a company that "fixes" these types of problems with publications?
- Google *engineered* this to be wrong so they could sell me on "fixing" these errors through their premium service?
- Google *imported* these errors from another source? Not *ResearchGate* anyway…

# Let’s Explore Some Examples

| TITLE | CITED BY | YEAR |
|---|---|---|
| Multiple algorithms for fraud detection<br>R Wheeler, S Aitken<br>Applications and Innovations in Intelligent Systems VII: Proceedings of ES99 ... | 227 | 2000 |
| A minimum dataset for a standard adult transthoracic echocardiogram: a guideline protocol from the British Society of Echocardiography<br>G Wharton, R Steeds, J Allen, H Phillips, R Jones, P Kanagala, G Lloyd, ...<br>Echo Research & Practice 2 (1), G9-G24 | 218 | 2015 |
| Mitral valve surgery plus concomitant atrial fibrillation ablation is superior to mitral valve surgery alone with an intensive rhythm control strategy<br>UO von Oppell, N Masani, P O'Callaghan, R Wheeler, G Dimitrakakis, ...<br>European journal of cardio-thoracic surgery 35 (4), 641-650 | 142 | 2009 |
| Diagnosis and assessment of dilated cardiomyopathy: a guideline protocol from the British Society of Echocardiography<br>T Mathew, L Williams, G Navaratnam, B Rana, R Wheeler, K Collins, ...<br>Echo Research & Practice 4 (2), G1-G13 | 102 | 2017 |
| Effect of transcatheter aortic valve implantation vs surgical aortic valve replacement on all-cause mortality in patients with aortic stenosis: a randomized clinical trial<br>T Fairbairn, I Kemp, A Young, C Ronayne, J Barton, J Crowe, L McQuade, ...<br>Jama 327 (19), 1875-1887 | 99 | 2022 |
| The role of echocardiography in guiding management in dilated cardiomyopathy<br>DE Thomas, R Wheeler, ZR Yousef, ND Masani<br>European Journal of Echocardiography 10 (8), iii15-iii21 | 97 | 2009 |
| A minimum dataset for a standard transoesophageal echocardiogram: a guideline protocol from the British Society of Echocardiography<br>R Wheeler, R Steeds, B Rana, G Wharton, N Smith, J Allen, J Chambers, ...<br>Echo Research & Practice 2 (4), G29-G45 | 41 | 2015 |
| Guidelines for transoesophageal echocardiographic probe cleaning and disinfection from the British Society of Echocardiography<br>P Kanagala, C Bradley, P Hoffman, RP Steeds<br>European journal of echocardiography 12 (10), i17-i23 | 38 | 2011 |
| Stress echocardiography in coronary artery disease: a practical guideline from the British Society of Echocardiography<br>RP Steeds, R Wheeler, S Bhattacharyya, J Reiken, P Nihoyannopoulos, ...<br>Echo Research & Practice 6 (2), G17-G33 | 35 | 2019 |
| Charge-coupled devices for the ESA Euclid M-class mission<br>J Endicott, S Darby, S Bowring, D Burt, T Eaton, A Grey, I Swindells, ...<br>High Energy, Optical, and Infrared Detectors for Astronomy V 8453, 19-26 | 33 | 2012 |
| The role of echocardiography in the management of atrial fibrillation<br>R Wheeler, ND Masani<br>European Journal of Echocardiography 12 (10), i33-i38 | 26 | 2011 |
| The Euclid VIS CCD detector design, development, and programme status<br>AD Short, D Barry, M Berthe, N Boudin, O Boulade, R Cole, M Cropper, ...<br>High Energy, Optical, and Infrared Detectors for Astronomy VI 9154, 234-246 | 23 | 2014 |
| A systematic approach to echocardiography in hypertrophic cardiomyopathy: a guideline protocol from the British Society of Echocardiography<br>N Smith, R Steeds, N Masani, J Sandoval, G Wharton, J Allen, ...<br>Echo Research & Practice 2 (1), G1-G7 | 17 | 2015 |
| A safety checklist for transoesophageal echocardiography from the British Society of Echocardiography and the Association of Cardiothoracic Anaesthetists | 13 | 2015 |

**Note how now that I have received the email**, and are aware of this vast mis-attribution, it would be unethical of me *not* to engage and try to fix it by joining their platform, uploading all of my publications, dis-attributing all of those now attributed to me, uploading my photo and CV, and on and on, and on.

*Whatever Google’s motivation, they have created a situation where I engage and work for free or I am seen as being unethical.* Including one of my actual publications as the first one is a beautiful, inspired touch of emotional manipulation.

# Let's Explore Some Examples

**Fourth case, medical ethics**: the designed system could detect whether someone was starting to suffer from cognitive decline disorders (such as Alzheimer's disease) from their physical motions and speech.

Is it ethical to train our systems and publish papers but not tell the subjects they might be falling seriously ill? What is our ethical obligation; who should we tell? What if the subjects did not approve use of their data in this way, and we found this out about them accidentally through normal ML and system testing?

Once a participant is found to be likely to be suffering from a cognitive decline disorder, should they be removed from the study because they can no longer give or understand *informed consent?*

# Let’s Explore Some Examples

**Fifth case, consent and ethical use of information**: early on, participants had to agree to use of their data, presence of the system in their homes, etc. But these were extremely poor people in rural Bulgaria who did not even have a mobile phone, nor know how to use a computer or the internet. Training and consent discussions were in Bulgarian, and everyone agreed to consent.


But I disagreed. In project discussions, I said I did not think these people could understand what fair and informed data use meant, nor what the impacts of an ethical breach might be. I explained that I thought these people agreed because they want to be helpful and because there was coffee and cake, and they were surrounded by people who also agreed. Do you think they were capable of giving consent?

*What the Bulgaria Red Cross said…trust, blood pressure cuffs, and nightly visits…*

# Duty of Care

**This brings us to a critical point in ethics: *duty of care*.**

There are many definitions (see readings) but for me, duty of care means the obligation you have to another person, animal, or ecology, conferred by their trust in you, by your authority, or by a position of responsibility.

It is sometimes the case that care givers decide to do something that might seem strictly unethical, but do so because they believe their duty to care is a larger ethical obligation.

This is important because when we design AI systems, they tend to be *trusted*, and they tend to be seen as *authoritative.* They have default *duty of care.*

# Duty of Care

**Once when I was working at the WHO, I had to make a difficult ethical decision alone as time did not allow discussions.**

I later told my boss, Peter Beales, that I had doubts that I had made the right decision. He said: “Was it difficult to make the decision? Did you deliberate a lot internally, and feel bad after you made the decision? Then you probably did it right.”

He understood that I was struggling with making difficult decisions about duty of care, what rules we are willing to break to do the right thing for people who have trusted us to make decisions. He understood the only really *wrong* way of doing this is to ignore it and do nothing. We cannot always make the right ethical decisions, but we can always *try*.

If you remember nothing else from these lectures, remember this: **we cannot always make the right ethical decisions, but we can always *try*.**

# An Aside.

**In an earlier lecture I noted how important it is that you work with people much better and smarter than yourself.** I learned more from Peter Beales at the WHO in 10 days about kindness, ethics, and duty of care than I would learn in the entire rest of my life so far. He also taught me the real nature of humility – that most of what we do in our lives does not matter in any way – only the ways in which we can make the lives of the people around us better really matters. Everything else is *vanity*.

**That's what working with great people does.** Without Peter Beales, I would not be able to speak on ethics today. And you cannot pay such people back, you have to pay it forward to others.

*Find your great people, who make you a better, smarter, more open, kind, and ethical person.*

# Let's Explore Some Examples

**One more, on ethical decision making**: the main SAAM system was built on accepted rules for diagnosing problems, etc., including in some "soft" medical contexts. It was entirely *declarative*, and every decision or recommendation could be explained in the context of best practices in the field, including declaring confidence measures in decisions.

Forgetting all of the legal, certification, and regulatory issues, should the system ever be allowed to work on its own without having a human in the loop for every decision or outcome? We calculated we could keep a human in the loop for 100 customers/clients/elderly homes. But to scale to 1 million people, it would be impossible. Should we deny them essential care and monitoring because we do not trust the system without a human in the loop? Is this an engineering and design problem, or a legal or a policy one?

# Learning

**In the SAAM project I learned a lot about ethics.** One thing I learned from SAAM social services partners was that they felt they were acting ethically, even if not always strictly legally, if they were *enhancing or preserving the dignity and quality of life of the people they interacted with and served*.

But who gets to make that decision and distinction? They felt *obligated* to make such decisions because they were in a position of *trust*. This definition of ethical behaviour, as rooted in common sense and our duty-to-care for others, our community, and our planet, links back to ancient human moral structures and has served very useful to me in my professional career and life.

If you act in a way that seems unethical or illegal, but are acting in *good faith* for the benefit of others, is it still unethical behaviour? What if you do nothing to avoid responsibility for a decision? What if you are acting in *bad faith*, are your actions still ethical?

# AI and Governance

## Wyoming voters face mayoral candidate who vows to let AI bot run government

'There's a new intelligence in town' as Victor Miller and his ChatGPT bot, Vic, plan to lead Cheyenne in a hybrid format


Victor Miller's customized ChatGPT bot, Vic, plan to run for mayor in Cheyenne, Wyoming. Photograph: Jaque Silva/Sopa Images/Rex/Shutterstock

AI and governance promises to be a difficult issue in the future. Should we trust the running of our society to a system that can't make the recipe for a pizza that won't kill us?

# Special Guest: Oskar Sandholt

Oskar is the City Manager for the City of Reykjavik and is responsible for many of Iceland's most critical government services, including it's digitalization and public services portals such as Island.is.  He is also trained in European ethics practices and laws.

# For Next Class

We have two weeks not yet scheduled with course content. I can try to customise these weeks to your interests. What subjects are most interesting or relevant for you?

Examples:

- Complexity and chaos theory and how it determines ethical design
- Knowability, explainability, trust
- Understanding search and state/decision spaces, and why they may determine ethical outcomes
- Understanding AI use in society, AAL, medicine, safety critical domains
- Why real AI people hate ChatGPT and find it inherently unethical
- Careers and ethics: CVs and resumes, AI as cheap entertainment
- Understanding the role of automation and ethical work including where AI displaces workers
- Deeper technical case studies and discussions including design details
- Understanding policy issues related to ethics and best practices
- An extended ask me anything about AI, ethics, systems design, etc.

# Ask Me Anything,

…and I mean *anything*.