AI and Ethics

Richard Wheeler

Reykjavik University

Something You Rarely Hear

Good scientists have good *answers*, great scientists have great *questions.* A good scientist excites you about *their* work, a great one excites you about *your own* work.

Intelligence is a bit like jazz – you have to surround yourself with people much better than yourself. If you work with smart people, you will become smart, if you work with stupid people, you will become stupider and stupider. Do not underestimate this effect.

I am not particularly *smart*, I am particularly curious, observant, and *fast moving.*

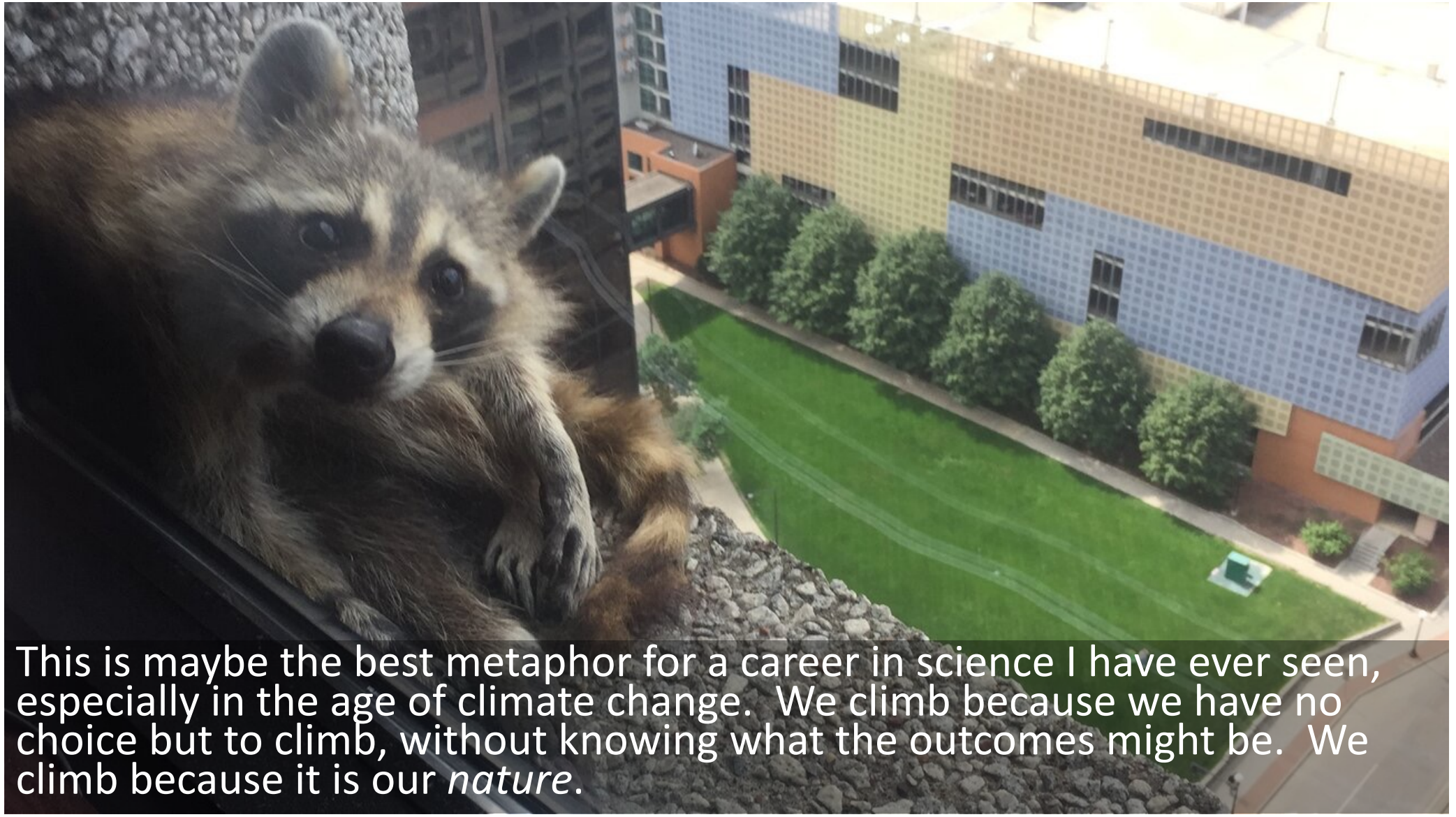
Those Who Try

…and history belongs to those who *try*.

Examples: University of London, World Health Organisation, EU Working Group on Teledemocracy, Arhus conference…

Ethical?

Those who try: *The Long Climb*

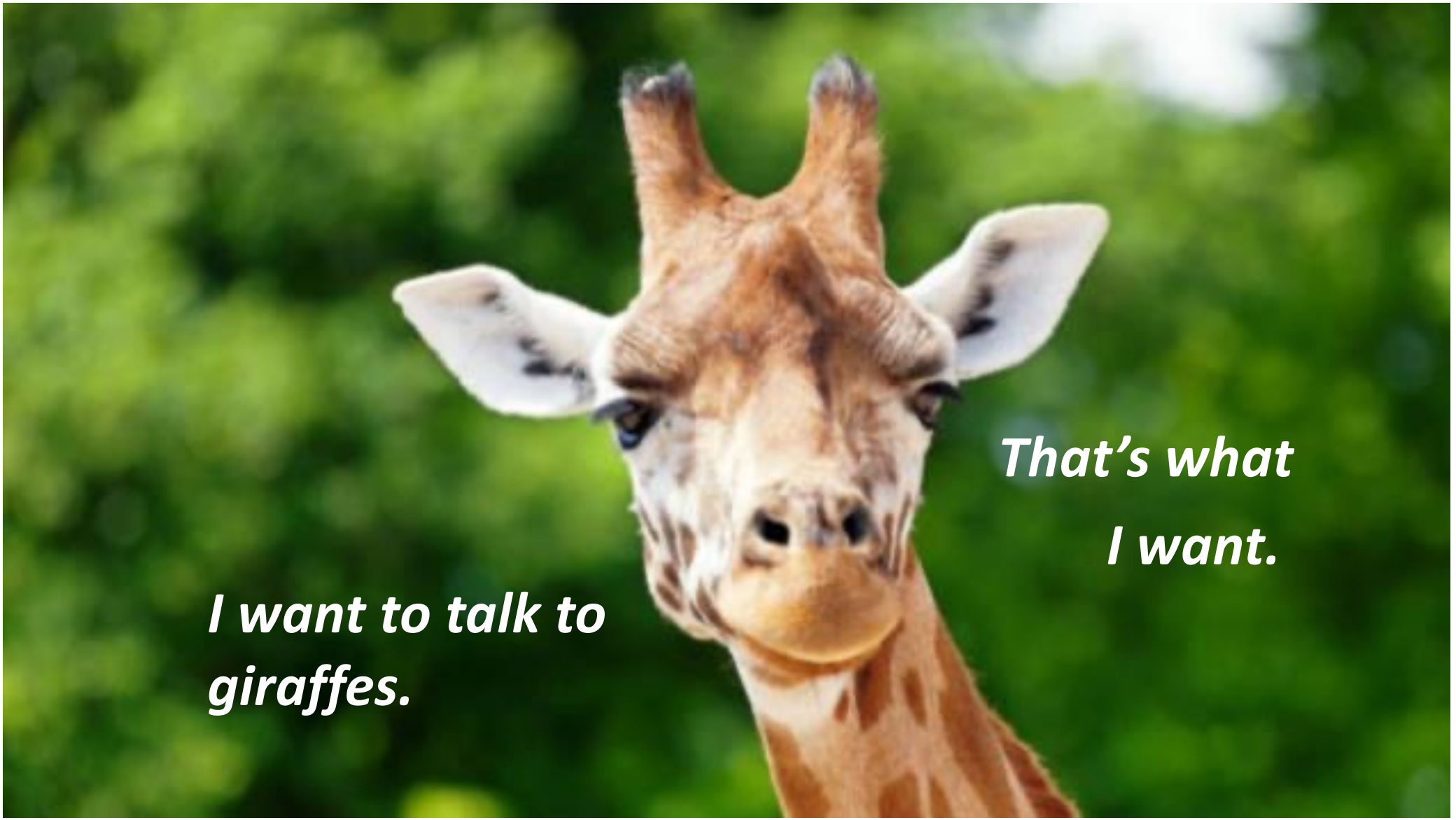

This is maybe the best metaphor for a career in science I have ever seen, especially in the age of climate change. We climb because we have no choice but to climb, without knowing what the outcomes might be. We climb because it is our *nature*.

Let’s Talk Ethics and System Design.

You will remember I asked you all to think of a use for your mobile phone that does not exist yet, and then we will present them together and discuss them to identify design and ethical problems.

Here’s mine.


I want to talk to
giraffes.


That's what
I want.
I want to talk to
giraffes.

How Animals Develop Regional Accents

Whales, bats, and birds have local dialects.

BY CLAUDIA GEIB JUNE 20, 2017

944

A singing Hermit Thrush (*Catharus guttatus*). MIRCEA COSTINA/ ALAMY

THE SONG OF THE HERMIT thrush sounds a bit like an orchestra warming up: the

How Giraffes Communicate

The giraffe (Giraffa camelopardalis) is the world's tallest mammal, standing as tall as 18 feet. They live in herds of anywhere from 5 to 20 giraffes. Within these herds, giraffes do communicate with one another, although they are often thought to be silent animals.

Humans cannot hear most of the communication between giraffes because they communicate infrasonically, with moans and grunts too low for humans to hear. Mother giraffes sometimes use whistles to warn or call their young.

Other ways giraffes communicate are with their eyes and by touching other giraffes in the herd.

Why do giraffes hum at night?

Despite their species' quiet reputation, giraffes at three zoos have been recorded humming to each other.

RUSSELL MCLENDON

September 22, 2015, 2:55 p.m.

7.1k Tweet 83

Giraffes declared an endangered species in Kenya

03 December 2018 - 07:29

BY ODWA MJO

Kenya has declared its giraffe population endangered.

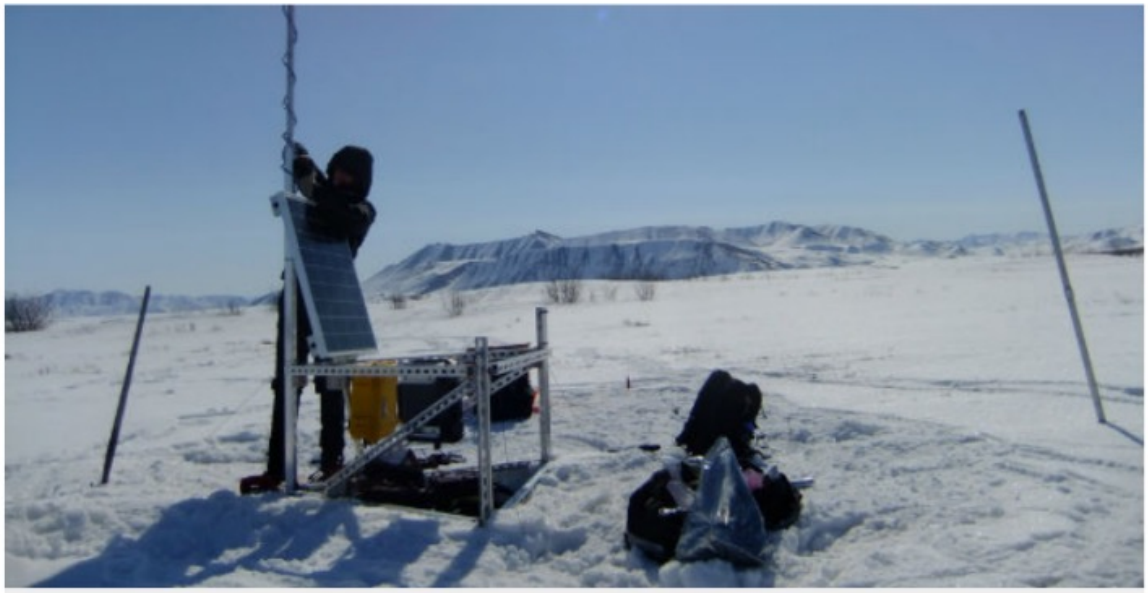
AI is being trained to recognize giraffes. Here's why

A.I. has been used to detect and recognize unique giraffe markings.

Researchers Develop an Artificial Intelligence to Analyze Birdsong in a Warming Arctic

BY KIM MARTINEAU | JUNE 20, 2018

11 4 Comments



Researchers set out four microphones like this one in the foothills of Alaska's Brooks Range to record songbird calls over five breeding seasons. In a new study, they describe an automated tool for sifting through thousands of hours of recordings to estimate when migratory birds arrive en masse in the Arctic. (Photo credit: Heather Greaves)

Proposal

- *Listening to giraffes will tell us a lot about regional migrations, group contacts, and ecological degradation patterns and flora and fauna pathways.*

- Background in machine learning, including signal processing. Exposure to many ecological issues, network in ecology and zoology. Knowledge of natural languages and emergent and context-free grammars.

Proposal

But how shall we do it? Any ideas?

Proposal

- Over two years, collect discrete sounds samples from national zoos and across Kenya via the *Giraffe Conservation Foundation*.
- We ask experts in the field who understand giraffe behaviour to try some simple tagging of utterances to intentions, objects, contexts, etc.
- To these sample and tagged pairs we apply supervised learning from other zoological and especially musical feature extraction literature (Humphrey et al., 2012. *"Moving beyond feature design: deep architectures and automatic feature learning in music informatics"* in International Society for Music Information Retrieval Conference.)

Proposal

- We can also use generative adversarial networks to try our hand at simple communication (within ethical boundaries) to assess the efficacy of our models (Yang et al., 2017 — *“MidiNet: A convolutional generative adversarial network for symbolic-domain music generation”* in International Society for Music Information Retrieval Conference.)
- We can use LLM methods to create a starting syntax, and then port all this over to Android phone for creating a real-time giraffe conversation device.
- We could then try it out in controlled conditions … like …*breakfast.*


Is this ethical? Why might it not be?
Can we design away the ethical problems?

Ethical Problems

- Does the system and it's use violate your personal ethics?
- Does the system violate laws and regulations about treatment of animals? Or zoo and animal reserve best practices?
- What if we say something that traumatises the animal? What if merely speaking to a human is a traumatic event?
- Do ethics only apply to humans? What about animals, the environment? Is there any such thing as ethical product testing?
- Does what we might learn outweigh the ethical problems?


Wait, no, we
did that one.

Let’s do something new.

Something *practical.*

My New Idea, Real Idea, Real Project Proposal

The problem: when you are in a wheelchair, and look on your phone to see if a building or site is disabled accessible, it sometimes says “yes” or has a wheelchair icon, but never *how* to access a building. Then you arrive there for dinner with friends, or visit the doctor, and are stuck outside until you give up.

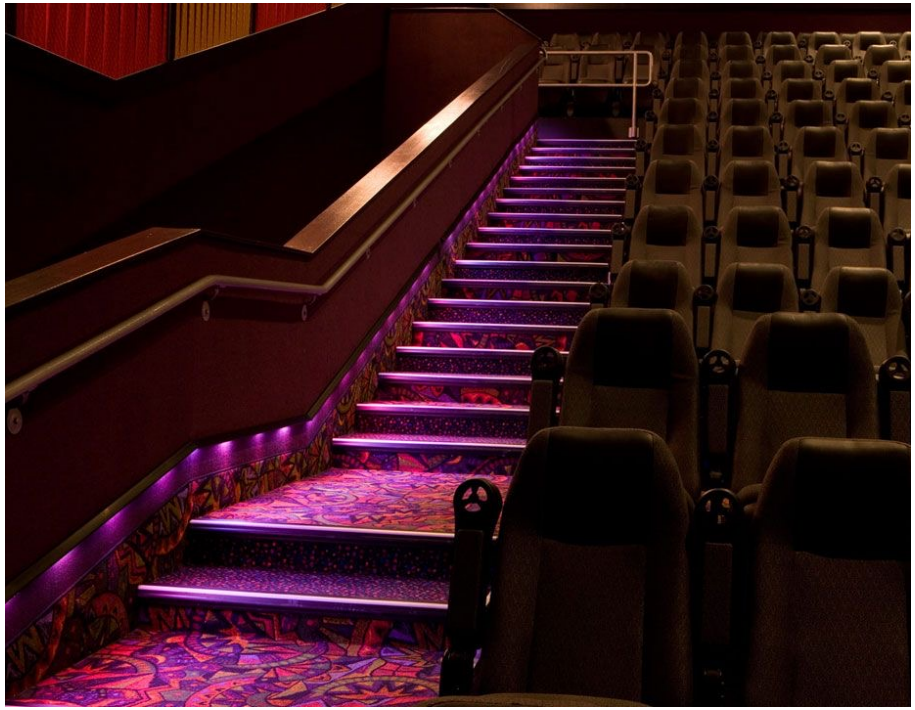
Disabled access is a legal, but primarily, an *ethical* issue.

Example

Reykjavik: Restaurant Forretta Barinn. Accessible? It says it is, but when you go there you see this (left) and give up. No dinner for you.

Example

Reykjavik: Movie Theatre Laugarásbíó. Accessible? It says it is, but when you go there you see this (left) and give up. No movie for you.



Example

Reykjavik: Floran Café in the Botanical Gardens. Accessible? It says it is, but when you go there you see this (left) and give up. No cafe for you.

What About Somewhere Actually Difficult?

What About Somewhere Actually Difficult?

Are there stairs at the end of this walkway to access the gates? Let's walk 2 kilometers and find out. :>(

Answer: yes, there were stairs at the end and no way down to the gates. Correct route? No idea.

What About Somewhere Actually Difficult?

Is there a route into this “accessible” art museum?
Or are we looking at it?

What About Somewhere Actually Difficult?

This café – accessible? It has a ramp, though, right?

Wheelchairs can’t cross the gravel in the parking lot, not even five metres, and cannot mount the ramp, it is too steep. On the other side of the door are stairs anyway.

University of Iceland

Lecture: they told me it was accessible. *Nope.*

I called them, sent a photo, eventually cancelled the lecture and went home.

Maybe I should have taken a Hopp scooter? Oh wait, I can't. :>(

Example

Why is it like this? Most locations are *technically* accessible if you knew what the right mechanism for access was. Even in my own building where I live, the main entrance is inaccessible. To access the building in a wheelchair, you'd need a key to the garage doors, go through there to the elevators, long path. But no visitor would know that, and the building is *strangely* not able to explain access to you. Why was this not the very first IoT application?

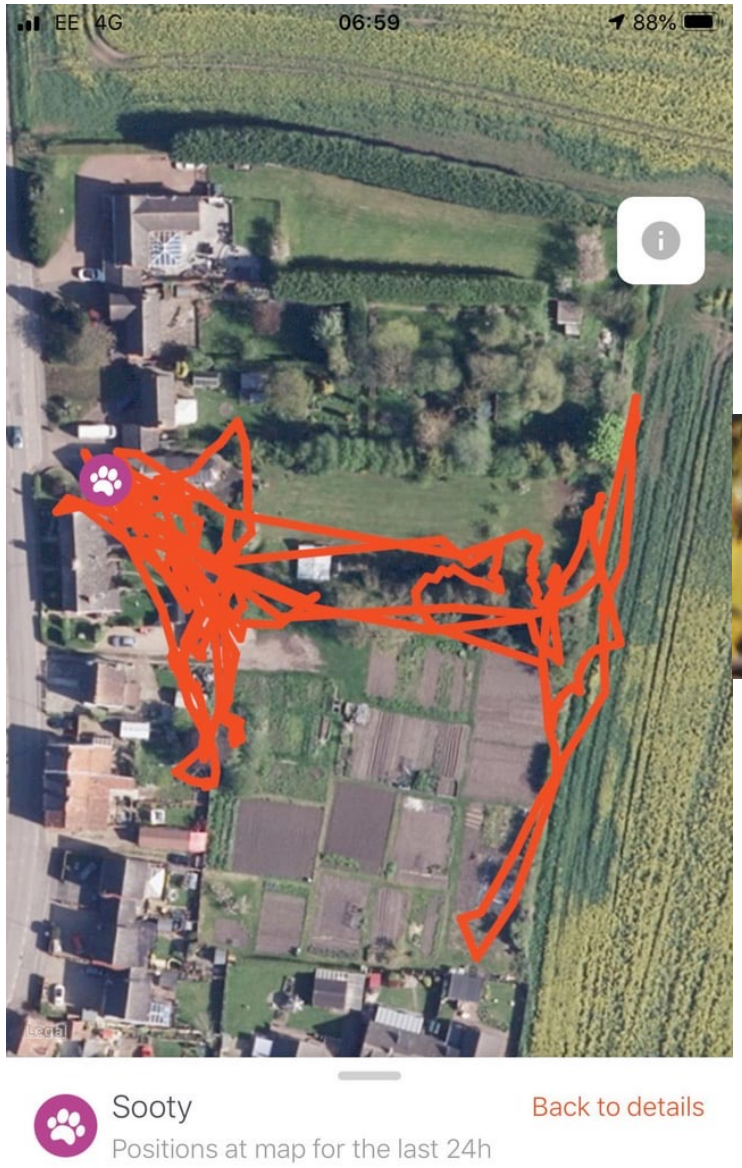
I want my mobile phone to show me actual routes that other wheelchair users have *successfully* taken to reach a specific location. (I also want my phone to handshake with the building, explain my disabilities, and load emergency and disabled access routes automatically, but that is a different project…)

Ideas?

Ideas for how to do it? Don’t bother looking it up online – there is *nothing*.

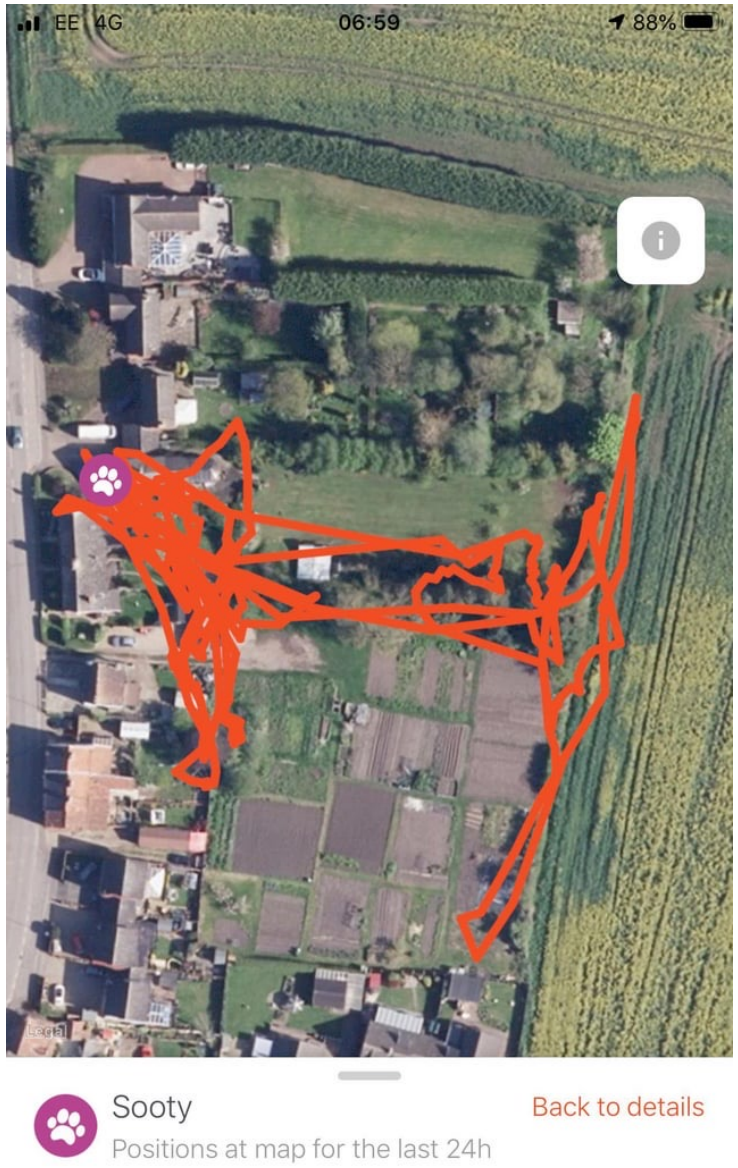
Come on, smart people, real problem, let’s hear solutions.

Okay, here’s my idea.



Idea

What do you see in this image of a pet tracking device's map (left)?



Idea

What do you see in this image of a pet tracking device?

I see where a cat *could go*. These are all valid routes for a cat, *physically,* proven by a single cat in a single day. Only a single cat needed to access these areas to prove they were … cat accessible. None of them were blocked or impossible for a cat.

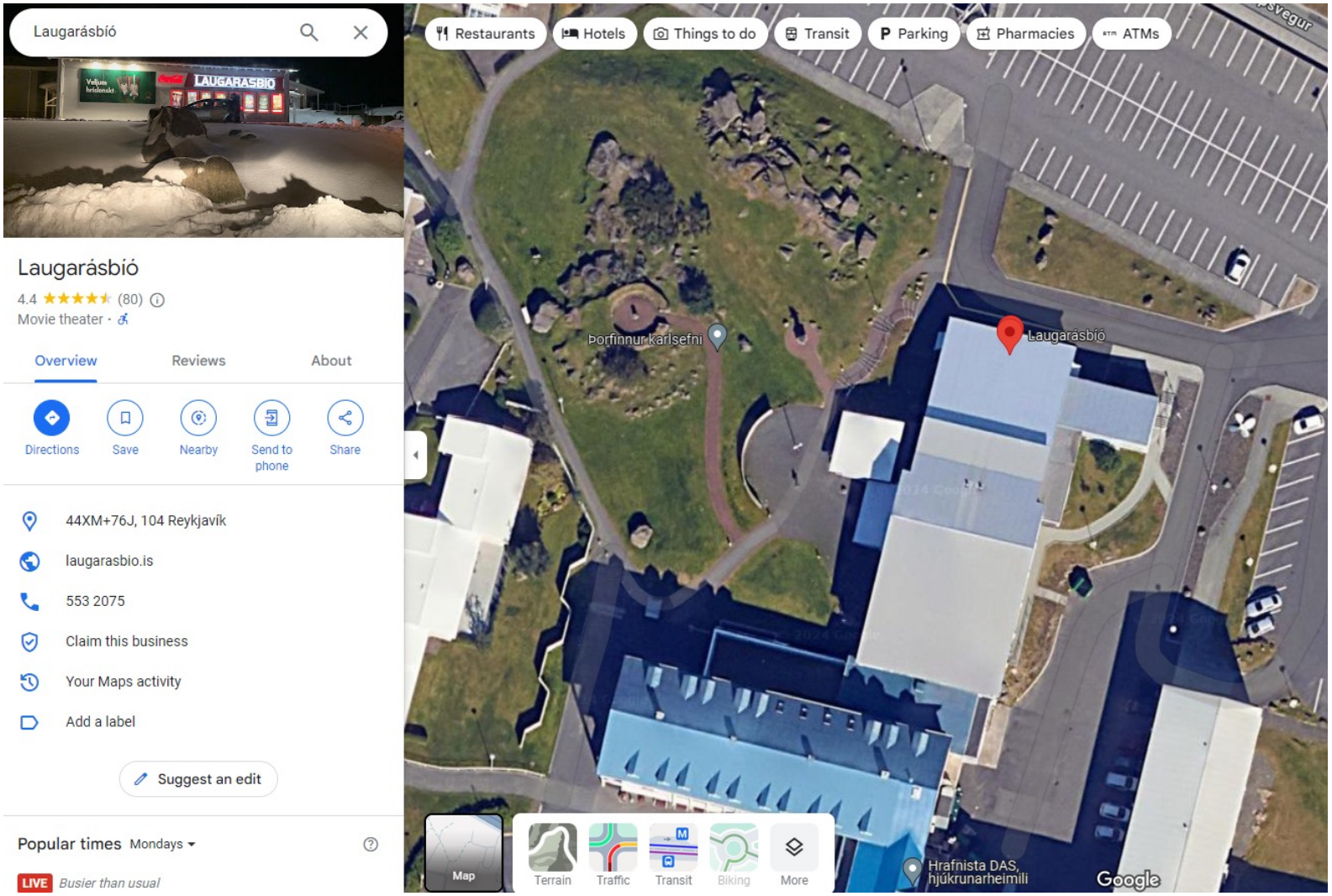
Got the idea yet?

Idea

My idea is to create a small program that you can install on your regular phone or an old unused mobile phone that has no interface, it just tracks you in your wheelchair to see where you *can go*. Charge it up, put it in the back pocket of your wheelchair, and it will create and upload a heatmap of where you've been. In this mode, it is *learning* new accessible routes.

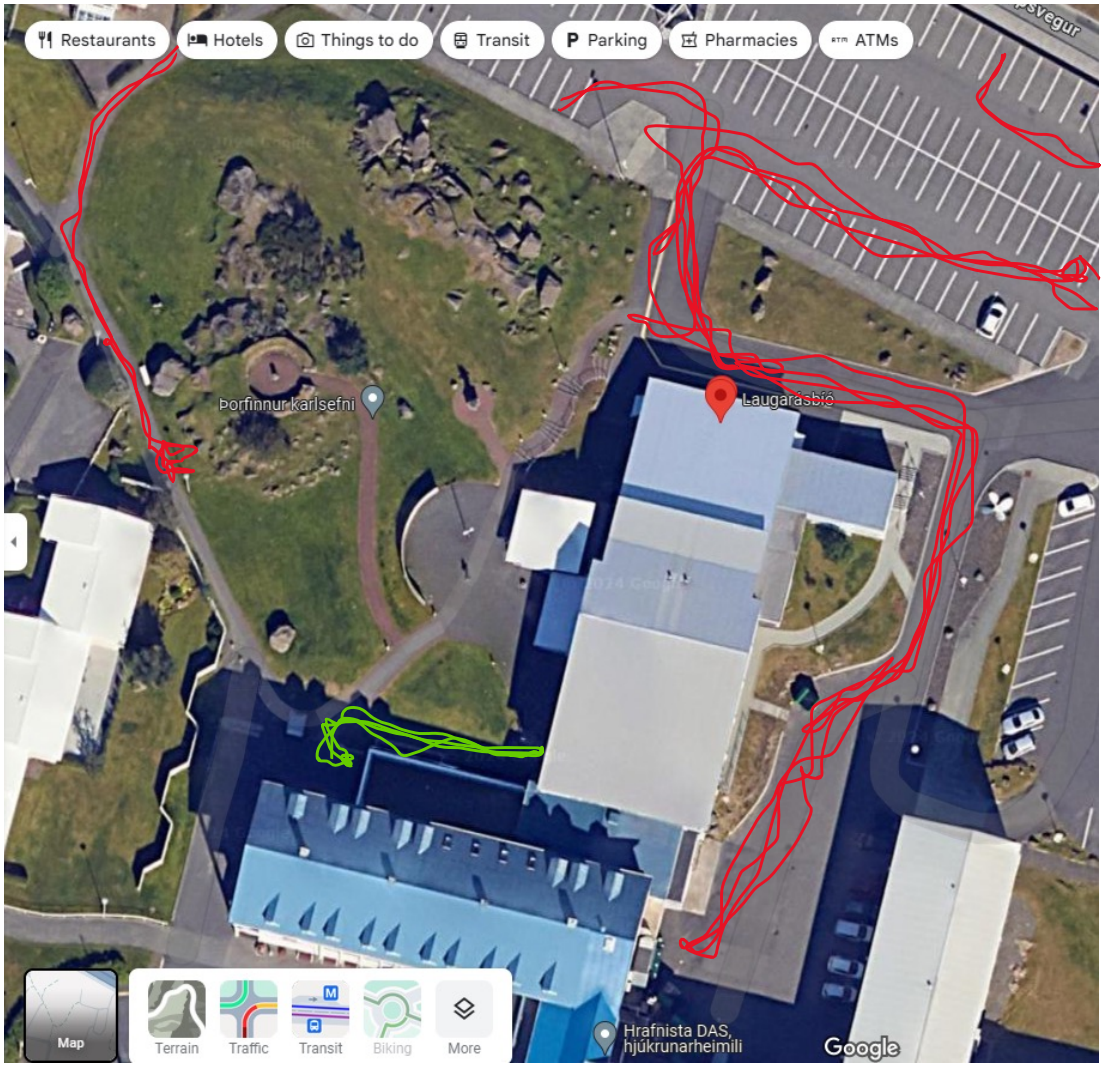
And then if you are accessing somewhere new, you can ask the system how other wheelchair users were able to access the café, the restaurant, the Botanical Gardens, the airport terminal, Perlan, Harpa, etc. Only *one user* needs to have accessed that place for this system to work. In this mode, it is *teaching* new accessible routes.

Back to the Movie Theatre



Information on Google maps – nearly none other than that it has a wheelchair icon.

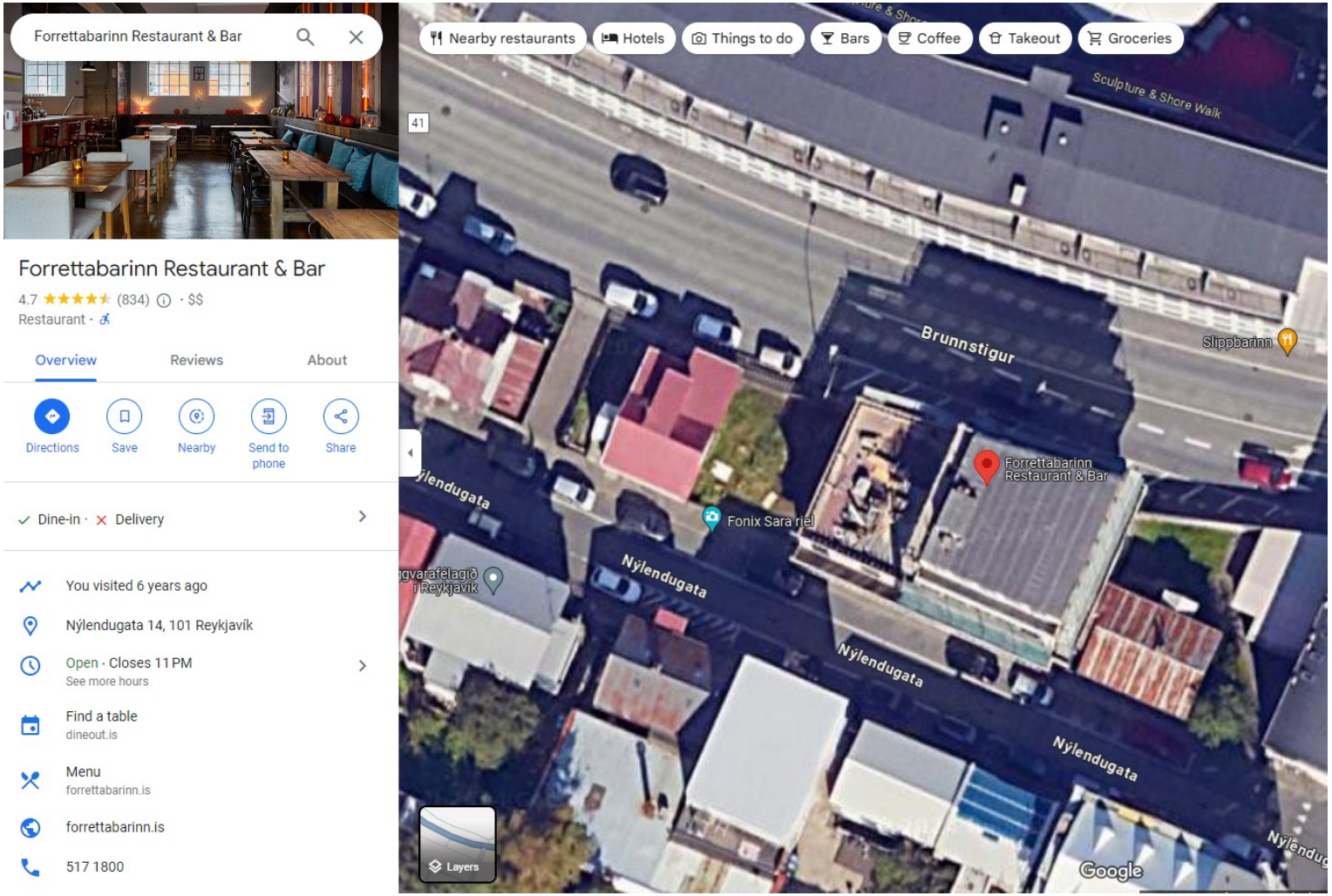
Back to the Movie Theatre: Richard's System



In red – failed journeys – no access, no movie.

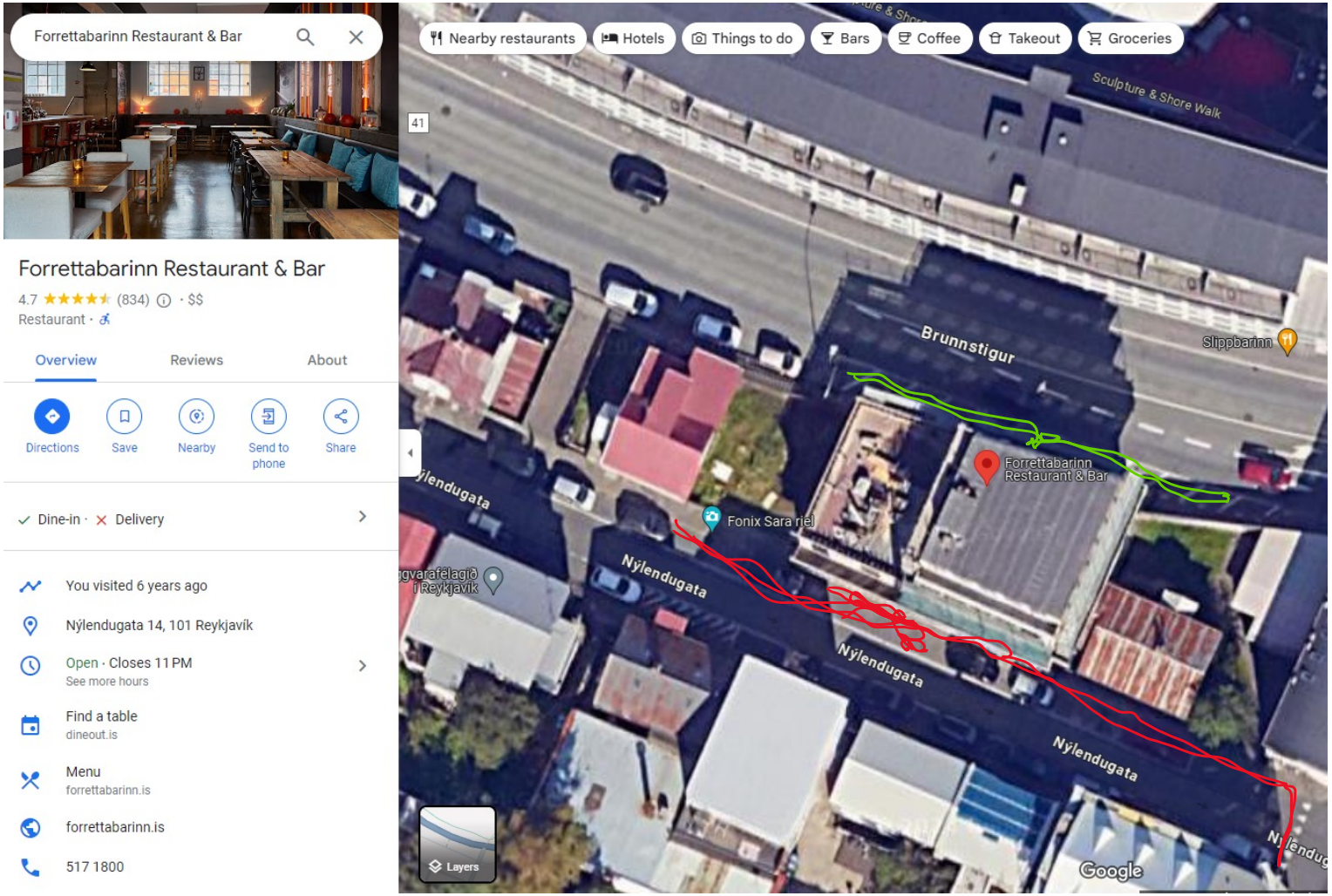
Green path – you park in the neighbouring lot behind the theatre, and access the theatre via a hidden door to the front near the screen.

Back to the Restaurant



Again, Google says it is wheelchair accessible, but as we've seen, it certainly is *not*.

Back to the Restaurant: Richard's System



Or is it? Green route: others have entered through the main entrance, so there must be some kind of hidden ramp. There is, in fact, a ramp they bring out and mount when you approach in a wheelchair.

Botanical Garden?

Don’t know, despite trying at least six times, we have never been able to access the café. Maybe there is a route? If you call them and ask they say there is a route but they cannot describe the path. :>(

The Application

Technical side – requires:

- An Android app that uses the GPS and OpenMaps to track routing, plus location detection to determine when a location is found.
- Simplified user interface to allow users to declare what type of disabled access they needed (wheelchair, flat path with sticks, no stairs, etc.) and a single button to push saying they got where they needed to go and confirm what it was.
- Server side to catch and overlay heatmaps, and to provide smart routing of green (successful) routes. APIs to OpenStreetMaps or similar to allow fast rendering and navigation integration of heat mapped layers.

Ethical?

Ethical issues with this idea?

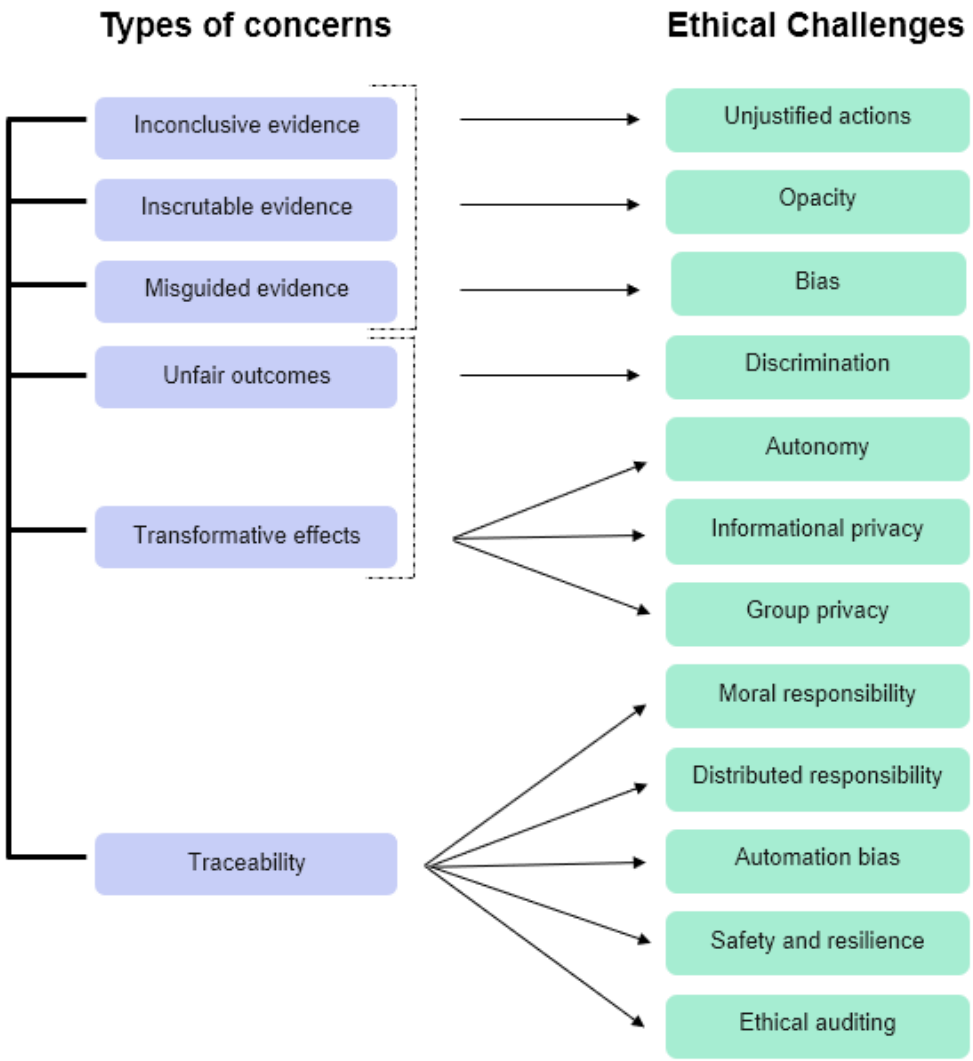
Design issues or ideas?

Your Ideas.

Now your ideas. They do not have to be good – we are not judging your mobile phone application, this is a *forum* exercise.

After each idea is explained, everyone should contribute a design idea, a risk, a possible ethical problem.

Might be helpful during the discussion: https://tinyurl.com/y7vjchzt



TYPES OF CONCERNS

- Inconclusive evidence
- Inscrutable evidence

When data are used as (or processed to produce) evidence for a conclusion, it is reasonable to expect that the connection between the data and the conclusion should be intelligible and open to scrutiny. Given the complexity and scale of many AI systems, intelligibility and scrutiny cannot be taken for granted. A lack of access to datasets and the inherent difficulty of mapping how the multitude of data and features considered by an AI system contribute to specific conclusions and outputs cause practical as well as principled limitations.

- Misguided evidence
- Unfair outcomes
- Transformative effects
- Traceability

ETHICAL CHALLENGES

- Unjustified actions
- Opacity
- Bias
- Discrimination
- Autonomy
- Informational privacy and group privacy
- Moral responsibility and distributed responsibility
- Automation bias
- Safety and resilience
- Ethical auditing

From: tinyurl.com/y7vjchzt

What Have We Learned

What have we learned in this exercise?

For Next Class

…each student will have about two minutes to present their idea, after which the rest of the class will give comments on design and possible ethical issues. The task is to identify and discuss ethical issues across a broad range of applied ideas. (done)

For Monday in four days, write and submit around one page of analysis of the discussion: what ideas you found most interesting, and discuss your own feeling for the ethical issues involved. The assignment is to show knowledge and insight into what constitutes ethical issues in AI systems. Explore what parts of the discussion you found most interesting given your background and scientific interests.

Ask Me Anything,

…and I mean *anything*.