Machine learning, social learning and self-driving cars

@JackStilgoe
Outline

- Responsible innovation as social learning
- Socialising machine learning
- Self-driving as a ‘solvable problem’
- Self-driving as misnomer
- Limits to responsiveness
Family hopes autopilot death leads to improvements

DRIVER WATCHING HARRY POTTER WHEN CRASH HAPPENED
US-27A (SR-500)

V02 Strikes Trailer of V01 and Goes Under

V01 Turning Left

V01 at FR

V02 Travels off Roadway and Strikes Fence

NE 140th Court

V02 Strikes Second Fence

V02 Strikes Power Pole

V02 Rotates to FR
A Tragic Loss

The Tesla Team  •  June 30, 2016

We learned yesterday evening that NHTSA is opening a preliminary evaluation into the performance of Autopilot during a recent fatal crash that occurred in a Model S. This is the first known fatality in just over 130 million miles where Autopilot was activated. Among all vehicles in the US, there is a fatality every 94 million miles. Worldwide, there is a fatality approximately every 60 million miles. It is important to emphasize that the NHTSA action is simply a preliminary evaluation to determine whether the system worked according to expectations.
“Perfect safety is really an impossible goal.”
Elon Musk, Sept 2016
TOUS COBAYES ?

Avec la voix de Philippe TORRETON
Librement adapté de l'ouvrage de Gilles-Éric SÉRALINI intitulé Tous cobayes ! Flammarion 2012
“Because the probability is so small, there is no practical possibility of determining this failure rate directly - i.e., by building, let us say, 1,000 reactors, operating them for 10,000 years and tabulating their operating histories”.

(Alvin Weinberg, 1972)

“Each nuclear power plant is its own test case. It may be categorized as an implicit experiment, and its most revealing case is the accident”.

(Krohn and Weingart, 1987)


(Wynne 1988)
### The *what*, the *how* and the *why* of innovation

<table>
<thead>
<tr>
<th>Products</th>
<th>Processes</th>
<th>Purposes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What are the likely risks and benefits?</td>
<td>• How should research and innovation take place?</td>
<td>• Why should this research be undertaken?</td>
</tr>
<tr>
<td>• How will the risks and benefits be distributed?</td>
<td>• How should standards be drawn up and applied?</td>
<td>• Who will benefit?</td>
</tr>
<tr>
<td>• What other impacts can we predict?</td>
<td>• How should risks and benefits be defined and measured?</td>
<td>• What are the alternatives?</td>
</tr>
<tr>
<td>• How might these change in the future?</td>
<td>• Who is in control?</td>
<td></td>
</tr>
<tr>
<td>• What don’t we know about?</td>
<td>• Who will take responsibility if things go wrong?</td>
<td></td>
</tr>
<tr>
<td>• What might we never know about?</td>
<td>• What if we are wrong?</td>
<td></td>
</tr>
</tbody>
</table>
Responsible innovation is ‘collective care for the future through the stewardship of innovation in the present’

(Stilgoe, Owen and Macnaghten 2013)
(also Stilgoe and Guston 2016)
Anticipation

Reflexivity

Inclusion

Responsiveness

Responsible innovation
Limits to responsiveness

- Inflexibility, irreversibility, commitment and other forms of lock-in
- Lack of reflexivity
- ‘Institutional void’ (Hajer)
- Presumptions of novelty act against learning (Rayner)
Approaches to social learning

1. Engineering – machines learn about society
2. Educational – learning happens socially
3. Institutional – social systems need to learn
   – Aircraft vs Nuclear power (Perrow)
4. Political – learning takes place in public
   – ‘Experimental government’ (Dewey et al)
Socialising
machine learning

diri noir avec banan
@jackyalcine

Google Photos, y'all f***ed up. My friend's not a gorilla.

Skyscrapers  Airplanes  Cars

Bikes  Gorillas  Graduation
What’s new?
Emerging Technology Hype Cycle

As of July 2015

Years to mainstream adoption:
- less than 2 years
- 2 to 5 years
- 5 to 10 years
- more than 10 years
- obsolete before plateau

gartner.com/SmarterWithGartner

© 2015 Gartner, Inc. and/or its affiliates. All rights reserved.
<table>
<thead>
<tr>
<th>SAE Level</th>
<th>Name</th>
<th>Narrative Definition</th>
<th>Execution of Steering and Acceleration/Deceleration</th>
<th>Monitoring of Driving Environment</th>
<th>Failback Performance of Dynamic Driving Task</th>
<th>System Capability (Driving Modes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No Automation</td>
<td>the full-time performance by the <em>human driver</em> of all aspects of the <em>dynamic driving task</em>, even when enhanced by warning or intervention systems</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Human driver</td>
<td>n/a</td>
</tr>
<tr>
<td>1</td>
<td>Driver Assistance</td>
<td>the <em>driving mode</em>-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the <em>human driver</em> perform all remaining aspects of the <em>dynamic driving task</em></td>
<td>Human driver and system</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>2</td>
<td>Partial Automation</td>
<td>the <em>driving mode</em>-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the <em>human driver</em> perform all remaining aspects of the <em>dynamic driving task</em></td>
<td>System</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
</tbody>
</table>

**Automated driving system ("system") monitors the driving environment**

| 3         | Conditional Automation | the *driving mode*-specific performance by an *automated driving system* of all aspects of the *dynamic driving task* with the expectation that the *human driver* will respond appropriately to a request to intervene | System                                             | System                           | Human driver                               | Some driving modes              |
| 4         | High Automation        | the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a request to intervene | System                                             | System                           | System                                     | Some driving modes              |
| 5         | Full Automation        | the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver* | System                                             | System                           | System                                     | All driving modes               |
“How quickly can we get this into people’s hands? If you read the papers, you see maybe it’s three years, maybe it’s thirty years. And I am here to tell you that honestly, it’s a bit of both,”

Chris Urmson (ex-Google), March 2016
‘Solving’ self-driving

AI IS THE SOLUTION TO SELF-DRIVING
If safety is the question, are self-driving cars the answer?

Source: The Economist, 2016
Seeing like a Tesla
Mobileye and Tesla part company

“We need to be there on all aspects of how the technology is being used, and not simply providing technology”

“This incident involved a laterally crossing vehicle, which current-generation AEB [automatic emergency braking] systems are not designed to actuate upon”

“Autonomous vehicles is an AI computing problem. It’s not a detection problem”

Mobileye

Nvidia
Modes of automotive learning

<table>
<thead>
<tr>
<th></th>
<th>Rule-following</th>
<th>Rule-making</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Autonomous vehicles’?</td>
<td>Technical and political entanglement</td>
<td>Technical and political ‘autonomy’</td>
</tr>
<tr>
<td>Modes of machine learning</td>
<td>Formal logic</td>
<td>End-to-end learning</td>
</tr>
<tr>
<td>Relevant Expertise</td>
<td>Domain expertise, ergonomics etc</td>
<td>Driving as a game</td>
</tr>
<tr>
<td>Pedagogy</td>
<td>Traditional</td>
<td>Autodidactic</td>
</tr>
<tr>
<td>Social constitution</td>
<td>Favours incumbents</td>
<td>Upstarts and start-ups</td>
</tr>
<tr>
<td>Strategy</td>
<td>Incremental innovation</td>
<td>Disruptive, permissionless innovation</td>
</tr>
<tr>
<td>Historical touchpoints</td>
<td>Automated highways</td>
<td>DARPA Grand Challenges</td>
</tr>
</tbody>
</table>
Escaping the risk frame

The social dilemma of autonomous vehicles

Jean-François Bonnefon¹, Azim Shariff²,³, Iyad Rahwan²,³

Fig. 1. Three traffic situations involving imminent unavoidable harm. The car must decide between (A) killing several pedestrians or one passerby, (B) killing one pedestrian or its own passenger, and (C) killing several pedestrians or its own passenger.
Machine control and social control
Alternative learning
Silence and Defeat the Tesla Autopilot Nanny Feature in V8.0
Mothra strikes back! Knocks me out of autopilot on a lonely stretch of road... (i.imgur.com)

submitted 9 months ago by Redebo MS P90DL + Mill Res holder

90 comments share
‘It's amazing technology... But Tesla shouldn't make their customers test pilots... [Autopilot] allows you to think it's working. And at a higher speed, you can't react. It's already hard enough to react when you are focused, that when you are not, and it errs, it's over.’

Mark Molthan, crashed August 2016
It's irresponsible to ship driving system that works 1,000 times and lulls false sense of safety, then... BAM!

Tesla Model S on Autopilot crashes into van parked on high...
It seems like some Tesla owners still aren't paying enough attention.
cnet.com
Human complacency is a serious but separate issue best addressed with education, monitoring & enforcement, not dumbed down safety systems.
Tesla crash report blames human error - this is a missed opportunity

Jack Stilgoe

In blaming human error for a self-driving car crash, US regulators have missed an opportunity to learn from such incidents.
The tyranny of urgency

We “cannot wait for perfect... We should be desperate for anything we can find to save people's lives,"

NHTSA Administrator Mark Rosekind

“If anybody legislatates against it, they should probably go to jail”

Engineer Alain Kornhauser

“Speed is of the essence here... We’re in a very important race against Asian and European countries that are moving forward with this technology”

Senator Gary Peters
‘Do the math’

‘If anyone bothered to do the math (obviously, you did not) they would realize that of the over 1M auto deaths per year worldwide, approximately half a million people would have been saved if the Tesla autopilot was universally available. Please, take 5 mins and do the bloody math before you write an article that misleads the public.’

Elon Musk, July 2016

‘If, in writing some article that's negative, you... dissuade people from using an autonomous vehicle, you're killing people’

Elon Musk, October 2016
Bounding the experiment?

- Broader environment, economy and society
- Landscape
- Infrastructure
- Other road users
- Other drivers
- Self-driving car and user
Questions for social learning: a research agenda

• Transitions
• Anticipating transport systems and car usage
  – What might be the pressures on machine readable infrastructure?
• Scale
• The governance of what?
  – Software, hardware, system?
Anticipatory governance
Histories

- Pre-history (Wetmore 2003; Vinsel, forthcoming)
  - Integrating vehicles and infrastructure
- Automation of car systems
  - Airbags
  - ABS/Traction control
  - Cruise control
  - ADAS
- 2008 → Machine learning
  - Neural networks and image recognition
  - GPUs
  - Data
The complexity of transitions

- Closed, homogenous systems are easy to imagine; Mixed, open systems are almost impossible
- The hand-off problem
- Smart tech can make people dumb
outline

• Tesla photo
  – Hardware and software – politics of novelty
  – It is a work in progress - learning in public – and innovating
• Crash
• Emerging sociotechnical system
• Responsiveness and social learning
• Imaginaries of AVs
• Governance
  – State as experimenter, not just laboratory
  – Safety – limits to responsiveness?
  – Data-sharing
    • Data is the currency of AV innovation
  – Conditions for experimentation
  – Software vs hardware – product liability
  – Democratising imagination
  – Who benefits? What are likely to be the transfers of power and money? Social constitution? Tech as legislation?
Perrow on the NTSB

• Why do aeroplanes get safer while nuclear power doesn’t?
  – Accidents are normal – not out-of-the-blue

• Learning as crucial
  – “past accidents also fail as warnings if the warning is available to only one part of the system” – and that was a problem with TMI
Starting with transport

- Infrastructure?
- Conditions for acceptability?
U.S. Air Carrier Safety Indices, 1920-1975
‘Never Neutral’
1960—The Radio Corporation of America and General Motors are proud to present the future of transport: cars that can drive themselves! You've heard about them for years, but now they're really here. Using sophisticated magnetic sensors and radio control, these cars steer autonomously, following guidance wires embedded in the road. Meanwhile, feel free to read, chat with your companions, or get some work done. Major highway systems could be in place as early as 1970, so we can promise that this technology will be coming very, very soon to a road near you.
Ethics and trolley problems
Anticipation
• Participatory, not predictive
• Understanding expectations, promises

Inclusion
• Public engagement
• User-driven innovation
• Value-centred design

Reflexivity
• 1st and 2nd order

Responsiveness
• Answering and reacting
• Understanding the political economy of innovation

Responsible (research and) Innovation
Anticipation

• From predictive to participatory
• Expectations and Imaginaries
• Tools
  • Anticipatory Governance
  • Vision assessment
  • Scenarios
• Barriers to anticipation
• Guston, 2012; van Lente, 1993;
• Fortun, 2005; Barben et al, 2008
Inclusion

• The ‘new’ scientific governance
• Dialogue and ‘mini-publics’
• The challenge of legitimacy
  • Input and outputs
• Wilson and Willis, 2004; Grove-White et al, 1997;
• Goodin and Dryzek, 2006; Irwin et al, 2013;
• Lovbrand et al 2011
Reflexivity

- From 1\textsuperscript{st} to 2\textsuperscript{nd} order
- Tools
  - Codes of conduct
  - Midstream Modulation
- Wynne, 1993; Schuurbiers, 2011;
- Swiestra, 2009; Fisher et al, 2006
Responsiveness

- Answering and reacting
- Diversity and resilience
- Value-sensitive design
- De facto governance
- Political economy of innovation
- Responsibility as metagovernance

Pellizoni, 2004; Collingridge, 1980; Friedman, 1996; Stirling, 2007; Kearnes and Rip, 2009
The rules of the road
New laboratory rules for self-driving cars?

- Software vs hardware
  - Product liability rules
- Ethical rules
  - German Ethics Commission
    1. property damage before personal injury.
    2. no classification of people by size, age etc.
    3. “If something happens, the manufacturer is liable”
- AI red flags
- Black-boxes
• Initial US rules. NHTSA: “The agency does not believe that self-driving vehicles are currently ready to be driven on public roads for purposes other than testing.”... until Tuesday
• Responsibility moving from state to federal level
  – 15-point voluntary checklist for manufacturers
Somewhere West of Laramie

SOMEWHERE west of Laramie there's a broncho-busting, steer-roping girl who knows what I'm talking about. She can tell what a sassy pony, that's a cross between greased lightning and the place where it hits, can do with eleven hundred pounds of steel and action when he's going high, wide and handsome.

The truth is—the Jordan Playboy was built for her.
Seeing like a Tesla
“If anything, Uber’s drivers are the R&D for Uber’s driverless future. They are spending their labor and capital investments (cars) on their own future unemployment.” – Douglas Rushkoff
Not sure. My wife has never had an accident, but when we take the car, I always insist on driving.

Autopilot would need to be very good indeed for me to trust it more than my wife.
• New California rules – talk in Berkeley about how misguided this move is...

• This could even be a separate paper – with Harry Surden?

• For now, the point is that things are framed in terms of safety only
  – Good to start with this – but as with Tesla story, ask what we can learn about tech going wrong for the governance of tech going right.
  – Safety frame is a huge barrier to social learning
  – We refuse to democratise the discussion of benefits

• The California DMV originally drafted regulations that demanded substantial collaboration between companies and the state on the terms and conduct of the social experiment.
• The ‘law lag’ is often actually a performance – ‘getting behind’ the technology, before it is defined
• A tightening of the safety frame
• Making deployment defined as use by members of the public) easier
• Accidents and Disengagements
  – (These requirements are still in the regulations – don’t understand the reporting)
• Made more permissive – innovation more Permissionless
  – ‘Embracing’ the technology actually means distancing oneself from it – disengaging
• Presumes technologies are defined as currently imagined – level 4 and 5, rather than a work-in-progress. Presumes technological completeness
  – Interesting definition of “operational design domain”, as opposed to real-world conditions
  – Interesting that training is required for AV drivers
  – Language of ‘accidents’ has been replaced by that of ‘collisions’
  – Data recorder included

• California announced in March 2017 their intention to relax these controls. Instead, they would put the onus on manufacturers to declare that their cars were safe and fully-insured and trust that the legal system would work out questions of liability and unintended consequences. The threat of being sued is imagined to be the strongest lever.
Social construction as performance
Mothra strikes back! Knocks me out of autopilot on a lonely stretch of road...
Redebo MS P90DL + Mill Res holder [S] 197 points 9 months ago

There I was, cruising along on AP doing 85mph on a lonely stretch of 93 between Kingman and Las Vegas, SUDDENLY, my driver console flashes red and commands me to take control of the vehicle. AP drops off. Cruise control drops off and I get the ominous warning, "Radar visibility has been reduced."

I see a lone gas station up the road and slow my roll into the fluorescent lighting of the canopy covering the gas pumps. I step out of the car while a car of German exchange students look quizzically at a MS parked next to a pump. Fearing the worst, I peek around the front of the car and was confronted with this sight.

The Demon Spawn of Mothra had attacked me and rendered my autonomy useless. Never fear though, a quick scrape with the window squeegee over the radar opening and my technology was restored!
And so ends another battle between Tesla and the forces of evil in the galaxy. :)

permalink embed
Escaping the risk frame

Fig. 1. Three traffic situations involving imminent unavoidable harm. The car must decide between (A) killing several pedestrians or one passerby, (B) killing one pedestrian or its own passenger, and (C) killing several pedestrians or its own passenger.
There is a blind spot in AI research

Fears about the future impacts of artificial intelligence are distracting researchers from the real risks of deployed systems, argue Kate Crawford and Ryan Calo.
Yet as with the robot apocalypse, the possibility of a driverless car weighing up ‘kill decisions’ presents a narrow frame for moral reasoning. The trolley problem offers little guidance on the wider social issues at hand: the value of a massive