The Swedish Law and Informatics Research Institute

Liability in the Era of AI

XXXV Nordic Conference
Stockholm

Stanley Greenstein, Jur. Dr.
Senior Lecturer/Researcher, dept. of law, SU
Stanley.Greenstein@juridicum.su.se

Background

- Conference theme:
  - Law in the era of AI
- This session's sub-theme:
  - To examine briefly the legal notion of liability/legal responsibility in the era of AI
  - To examine the effect of technology on law/assess the ability of traditional legal concepts to remain relevant in the era if AI

To what extent can we apply traditional notions of *liability* to AI agents?

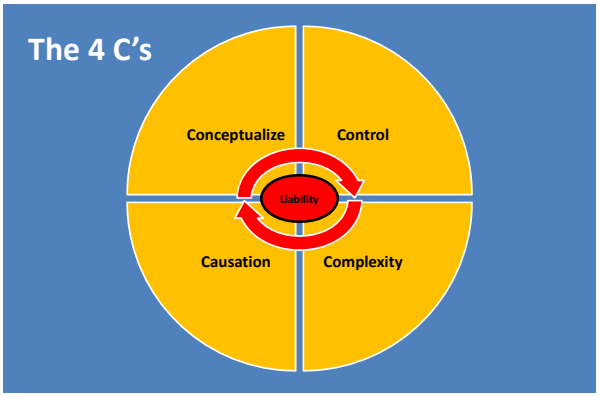
The vast legal landscape…

- Examining a concept such as 'liability' is challenging!
- Different legal traditions – civil law vs. common law
- Different areas of law – contract law, tort law, strict liability, legislation e.g. directives and regulations
- Criminal law or civil claims
- Military context?
- Some developments have received extra attention e.g. autonomous vehicles
- Some professions have different rules determining liability
- Legal context!
- …

Language

- Responsibility
  - A moral duty
- Liability
  - 'the state of being legally responsible for something : the state of being liable for something … something (such as the payment of money) for which a person or business is legally responsible … someone or something that causes problems' (Merriam-Webster legal dictionary)

The 4 C's



Conceptualize

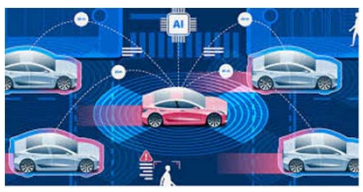
An issue I keep on coming back to…

What is AI?

Conceptualizing AI

Conceptualizing AI

Conceptualizing AI



Conceptualizing AI

Artificial Intelligence…where are we?

Autonomous robots?

'Reactive machines'
Simple input/output actions
Mere tools

Systems that:
1) think like humans
2) act like humans
3) think rationally
4) act rationally
Russell & Norvig

AGI/Strong AI/Singularity

Control

Control

- Bryant Walker Smith takes a linguistic approach in examining the term control
- Conventional control theory: systems are designed to achieve particular goals
- 'a goal-oriented action by a subject upon an object' (Zdzislaw Bubnicki)
  - External designer
  - Imposing external goals
- Works for clearly defined systems but confusing where the system boundaries not clearly defined
  - Humans: 'in control', 'in the loop', 'out of control', 'out of the loop'
  - Automated systems: 'under control', 'under human control', 'under computer control', 'out of control'
- Relationship between the notions of control and responsibility/liability
- Autonomous vehicles: who is responsible?
  - Owner/operator/seller/manufacturer/other party

Control

- Should we hand over control of certain decision-making processes to intelligent systems (e.g. robots)?
- If yes, then how do we structure the legal frameworks that determine legal responsibility/liability?
- Risk when AI gets it wrong:
  - 'Well, we new the risks'
  - 'Well, we knew that this is a highly speculative instrument'
- Unlikely that regular product liability rules will suffice
  - OK for 'tools' but not 'upper end of the spectrum AI'

Causation

Causation

- 'Causation is generally an application of a general principle to a special case for a special purpose' (Jan Hellner, *Causality and Causation in Law, Scandinavian Studies in Law*).
- It is used to:
  - To explain the occurrence of particular events
  - To predict future events
  - To control events
  - Too attribute moral responsibility and legal liability
  - To fulfil certain technical applications of physical theory
- Causal chains are important:
  - Event A is a cause of B and B is a cause of C
  - Does not mean we can accept the combined link A to C

Causation

- Essential for claiming liability is the link between the victims harm and the defendant's sphere
- The victim must prove that the damage originated from the defendants conduct or risk attributable to the defendant
- Victim must bring evidence
- However, difficulties to succeed where:
  - Less evident sequence of events
  - More complex interplay between factors (jointly/separately)
  - More links in the control of the defendant
- Standard of proof
- Explicability (complexity) … e.g. to measure success of claim

Causation

- A complication factor is the immense amounts of external input data and the difficulty establishing causation
  - Was the damage caused by a single original cause or by the interplay of multiple causes?
- A standard response:
  - No-one is liable
  - All the parties are jointly and severally liable (current majority view in Europe)
- Notion of linear causation giving way to complex, nonlinear interactions
  - Notion of 'combinatorial explosions (that) can outwit people's best efforts at predicting and mitigating trouble' (Dekker, 2009)

Causation

- 'No surgery can separate these inextricably entwined causes. No judge can isolate the "legal" cause of injury from the pervasive electronic hum in which they operate, nor separate causes from the digital universe which gives them their mutable shape and shifting sense. The result is a snarled tangle of cause and effect as impossible to sequester as the winds of the air, or the currents of the ocean. The law may realize that networks of intelligent agents are not mysterious black boxes, but rather are purposeful, artificial constructs. But that will not solve the problem of legal liability. The central doctrine of proximate cause, essential in the sorting out of multiple causes and tagging some in accordance with public policy, is useless when causes cannot be sorted out in the first place. '

(Karnow, *Liability for Distributed Artificial Intelligences*)

Complexity

Technology inspired by nature: Evolutionary Computing

- Goal of AI: to developing computer programmes with intelligence + the ability to self-replicate + learn + control their environment > looked to the natural systems > look at the human brain, human learning, evolution
- Pioneers of AI interested in biology + psychology
- Biology inspired areas of AI > neural networks, machine learning & evolutionary computing (GA's)
- 50's and 60's evolution studied as an optimization tool > idea to evolve a population of candidate solutions to a given problem using operators inspired by natural genetic variation and natural selection ('survival of the fittest')
- 4 requirements:
  - Fitness: ability to survive and reproduce
  - Variation: inherited or environment
  - Reproduction: creating new offspring
  - Heredity: parents pass on traits to offsprings

Genetic Algorithms

- GA a subfield of Evolutionary Computing
- In 1975 John H. Holland presents GA's in *Adaption in Natural and Artificial Systems*
- 'software that made software', code produced 'that no human would ever have written'
- 'Evolutionary robotics' – code designed by code within allocated constraints e.g. task and environment
- Population of 'chromosomes' (string of 1's and 0's)
- Creating a new population by 'natural selection'
- Genetic inspired operators of 'crossover', 'mutation' and 'inversion'
- Considered a strong tool for complex problem-solving:
  - Can easily adapt to an evolving environment – just like humans!
  - Innovative and can perform tasks that are too complex for manual programming
  - Ability to evaluate different possibilities at the same time

Genetic Algorithms

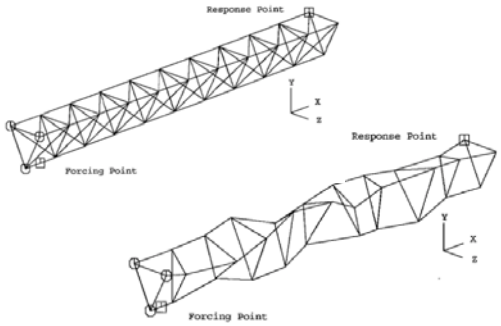
- So how does it work?
- 'Individuals of a population contribute their genetic material (called the genotype) proportional to their suitability of their expressed genome (called their phenotype) to their environment, in the form of offspring. The next generation is created through a process of mating that involves recombination of two individuals genomes in the population with the introduction of random copying errors (called mutation). This iterative process may result in an improved adaptive-fit between the phenotype of individuals in a population and the environment' (Brownlee, *Clever Algorithms*)

Complexity

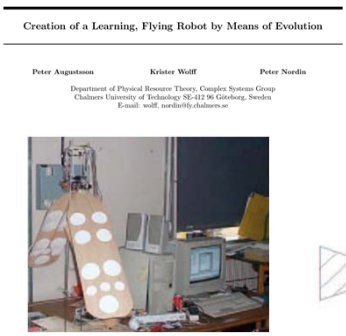
'We have neither pre-designed the behaviors of the robot, nor have we intervened during evolution. The robot itself and alone has developed … a set of strategies and behaviors as a result of the adaption to the environment and its own body… it is difficult to predict the robot behavior, due to the non-linearities and the feedback connections exploited for optimal navigation and obstacle avoidance'

(D. Floreano et al, Evolution of Adaptive Behavior in Robots by Means of Darwinian Selection)

Industrial design

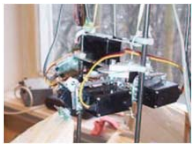


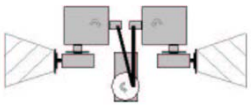
'Teaching' a robot to fly


Creation of a Learning, Flying Robot by Means of Evolution

Peter Augustsson Krister Wolff Peter Nordin

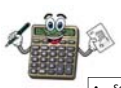
Department of Physical Resource Theory, Complex Systems Group
Chalmers University of Technology SE-412 96 Göteborg, Sweden
E-mail: wolff, nordin@fy.chalmers.se





Autonomous robots

- A future where:
  - Products of feedback loops
  - Robots given the instructions for the ultimate goals & to determine for themselves the means for achieving these
  - These means unpredictable by the original programmer or by the owner-operator
  - Software teaches itself by trial-and-error



- Software interaction measured and controlled
- Limited sources of unpredictability
- Disembodies software

- Operate in the physical environment
- Exposed to varied types of inputs (speed)
- Pace of the real world cannot be dismissed
- 'Highly porous' (highly integrated with their environment)

Challenges

European Commission White Paper on AI (Brussels, 19.2.2020 COM (2020) 65 final)

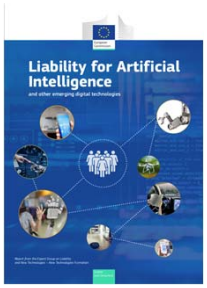
https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Legal challenges

- EU Commission White Paper
  - Safety legislation applies to products and not services (AI a product or service?)
  - Stand-alone software products covered by safety legislation?
  - Phrase 'placing on the market'? (when does this occur? continual updates?)
  - Long and complex supply chains – product liability law attached liability on the producer that places the product on the market (who is the producer in the long supply chains?)

Challenges

- EU Commission: *Liability for Artificial Intelligence and other Emerging Digital Technologies*
- Report from the Expert Group on Liability and New Technologies – New Technologies Formation (an expert group set up by the European Commission)



The EU perspective (NTF)

- AI technologies may possible cause harm (e.g. bodily injuries)
- Damage may still occur and the victims may be required to seek compensation
- In most Member State's this is usually done either by means of tort law (private law) & possibly in conjunction with insurance
- Tort law largely *unharmonized*
  - Exception: product liability law (Directive 85/374/EC)
- Strict liability is of producers for defective products is regulated by the Product Liability Directive (the rest by the Member States themselves)

Conclusions (NTF)

- Adequacy of existing liability rules is questionable
  - Formulated decades ago
  - Incorporate a monocausal model of inflicting harm
- The characteristics of new technologies is affecting the ability of victims to claim compensation
- Characteristics:
  - Complexity
  - Modification (updates, self-learning)
  - Limited predictability
  - Vulnerability (cyber security threats)
- Remedy: adjustments to national liability regimes

Some suggestions (NTF)

- Some of the suggestions:
  - High risk technology should be subject to strict liability of those in control
  - Where a service provider has a higher degree of control than the owner/user > this should be taken into account
  - Manufacturers of products to be liable for damage caused by defects in their products even where the defect was caused by changes made to the product under the producer's control after it had been placed on the market
  - Compulsory liability insurance
  - Victims should be entitled to the facilitation of proof (complex technologies)
  - Logging features
  - Destruction of data = damage
  - No need to give autonomous AI agents a separate legal personality status

Specific challenges (NTF)

- The more complex the digital technology becomes, the more difficult is will become to apply liability frameworks
- Causation (addressed above)
- Duty of care:
  - Usually the requirement to 1) identify the duty of care that should have been upheld and 2) prove that this was not the case
- Current legal regimes built upon the notion of human beings doing harm and monocausal harm
- Alteration of the initial algorithm + self-learning capabilities (autonomy)
- Forseeability: 'many systems are designed to not only respond to pre-defined stimuli, but to identify and classify new ones and link them to a self-chosen corresponding reaction that has not been pre-programmed as such'
- Internal workings of the original intelligent components may not be explicable.

USA

- 'Liability'
  - Contract, tort law & statutes
- Tort law
  - Negligence
  - Strict liability
    - 'liability no matter what'
    - Liability w/out negligence or other fault > usually applied to products
    - Consumer can sue whether or not the maker exercised 'due care' or was negligent
    - Design defect, manufacturing defect, ultrahazardous activity & failure to warn

Forseeability

- Common denominator = foreseeability
- A type of predictable harm to a predictable group of potential victim
- 'Negligence and strict liability were born and raised in a Newtonian universe, the universe of billiard balls hitting billiard balls, car hitting cars; force, mass and reaction; and machinery executing one step at the time. The risk discernible in this world are the consequences of Newtonian mechanics, which is linear: A causing B causing C … With autonomous robots that are complex machines, ever more complex as they interact seamless, porously, with the larger environment, linear causation gives way to complex, nonlinear interactions … the problem is not ignorance; the problem is the limits of knowledge' (Karnow).

Assessing liability IRL

- High end expert robot produces a bad decision in the hospital context
- Example: liability where a high-end expert system in a hospital makes a bad decision?
  - Human decision-making error: call experts to give evidence concerning how and why a decision was made and whether it was a sound decision – a judge to determine eventual breach of the appropriate standard of care
  - What is the 'appropriate standard of care' for the expert system?
    - No previous cases to base an opinion on
    - What would the 'reasonable robot' have done?
  - How do you get the robot to explain its decision when the technology lacks explicability?

Millar & Kerr: *Delegation, relinquishment, and responsibility: The prospect of expert robots*

To what extent can we apply traditional notions of *liability* to AI agents?