Ethiek – HC 6, 16-12-2020, Henry Prakken (English lecture)

Part 1 Introduction AI & Law
The topic of this lecture is artificial intelligence and law.

Some press on AI & law
Cnbc.com → Lawyers could be the next profession to be replaced by computers.
Nytimes.com → AI is doing legal work. But it won't replace lawyers yet.
Bbc.com → the robot lawyers are here – and they're winning
Daily Telegraph → "**Artificially Intelligent 'judge' developed which can predict court verdicts with 79% accuracy**" (…) "Computer scientists … developed an algorithm which can not only weigh up legal evidence, but also moral considerations."

The ECHR predictor
So, what is this system that can predict court verdicts with 79% accuracy? It is actually a research project on an algorithm that predicts outcomes of the European court of human rights. It is a machine learning application and its trained-on training data and tested on test data. Its trainings data was the full text of a large collection of decisions of the court concerning three articles in the European convention of human rights. And the **task** of the algorithm was to predict whether the court ruled that particular article was violated. In **79%** of the cases the ECHR predictor answered correctly. But this is by no means break true. First of all, you should know that an algorithm is a **yes/no question**. So, you have 50% chance. So, you should compare 79% to 50%. Moreover, it is very easy to score even higher because the human court finds violation for these three articles in 84% of the cases. The system **does not predict outcomes** because it needs most of the decision to be predicted. The last shortcoming of the system is that it **cannot explain** its answers. It has no understanding of the text; all it does is counting the frequencies of word combinations.

Follow-up research
The point that it cannot explain outcomes was illustrated in follow-up research. They studied the most relevant word combinations. For 'violation' it was 'death penalty', that the applicant, the public prosecutor. This gives no legally meaningful information why there would be a violation. For 'no violation' it was even worse, the word combinations were the first applicant, district prosecutor office, the urus martan.

Predicting outcomes of SCOTUS decisions
Returning to the first fundamental limitation that the system does not predict because it already needs most of the decision to be predicted. This is different in another research project. This project was meant to predict outcomes of the US Supreme court cases. This was a binary decision because the algorithm had to predict if the Supreme court would overturn the decision of the lower court yes or no. It did on the bases of a database on SCOTUS. The data contains mostly not related to the merits of the case. Like which presidents appointed the judges, personal data about the judges and trends in SCOTUS decisions. This system predicted 70% of the decisions correctly. But again, you have to compare it with 50% because it is a yes/no answer. Although it really predicts outcomes, the second limitation still exist because most data are not about the merits of the case and it also cannot explain its prediction in legal meaningful terms. And that is a shortcoming if you want to use this system for deciding cases.

Prediction is not decision making
There is a more fundamental reason why such case outcome predictors are unsuitable for modelling legal decision making, because judges don't predict but **decide**. They try to justify their decisions on **legal grounds**. The predictive algorithms don't do this at all. They have no way to justify their decisions on legal grounds because they don't understand the text. Moreover, how can we know that a predicted decision is correct if it cannot be explained on legal grounds. Judges think that it might happen soon, that judges who deviate from outcome predictors might have to explain why they deviate. Such a deviation could only be accepted in the case if there are special circumstances. Otherwise, they would

have to follow their predictions. This would not make any sense because in no way we can say that an algorithm prediction is a legally correct one. Can there maybe be a AI-system that is knowledge-based?

Part 2 History of AI (& Law)

Institutional history
Many people think that AI is the synonymous to machine learning, but this is not the case. Many people also think that the field of AI and law is quite recent. But that is also not true at all, it goes back to 1980's. We had Florence conferences in 1982, 1985 and 1989 and ICAIL conferences since 1987, JURIX conferences since 1988 and AI & law journal since 1992. We also have two landmark papers: Taxman and British Nationality Act.

Symbolic vs. data centric
Throughout the history of AI, you see that two main approaches have been pursued. The **symbolic/ cognitive approach** to AI was prevalent in the early days. The attempt was to program explicit models of aspects of human cognition. The advantage is transparency, they can easily point at which knowledge they applied and how they applied it. But a main disadvantage is that it turned out to be very hard to input the knowledge needed for realistic problem solving into a computer understanding form. This is very hard to model symbolically. The other approach is the **statistical, data-centric** AI. The computer automatically learns interesting knowledge from large patterns of data without being told explicitly what it needs to know. This had been successful in many areas. A big disadvantage is that the machine learning algorithm tend to be very hard to understand for humans. These systems cannot easily give explanations. Often these approaches have to be combined to make things work.

Some history on cognitive, symbolic AI
In the first years people still did try to model **general** intelligence. But this is too ambitious. Around 1970 the idea arose of so-called **expert systems**, later called **knowledge-based systems**. People said let us limit the problem domain to which the system has to be applied to the expertise of a human expert in a limited problem like diagnoses and treatment of infections. That led to some successes, the MYCIN system was a very influential project, it shown to perform better than human doctors, but it was never applied in practice. In general, the knowledge accusation bottleneck remains the problem to turned out to be a general quite hard to get the knowledge out of the human expert brain into the computer. Human experts don't make all their knowledge explicit. In the law some computer scientist thought things are different, their knowledge is explicit. You can model the legal rules in logical format.

Three levels in legal decision making
This mechanical deductive view on what is legal reasoning just rule application is far to simplistic. If you look at a legal case, there are three different decisions that have to be made. First **determining the facts of the case** (legal proof). This is not rule based. When the facts have been determined they have to be **classified** under the conditions of a statutory rule (legal interpretation). Third, you have to **apply** the statutory rule. Even then there is room for exceptions.
Modelling legal reasoning – deductive rule-based systems

Legal knowledge-based systems in practice
Still while this **rule-based** view on legal reasoning is in general to simplistic, one of the most successful AI-systems is based on this rule-based view. Not so much in court, but in public administration. The majority of the decisions has some legal defect, so this is not very good. Governments introduced rule-based systems which don't really automate legal reasoning but automate the logic of regulations. In the early days the facts still had to be decided by the humans. But currently, the systems are fully automated and there is no human intervention anymore. The facts are often taken from case files and government databases. If there is still a need for interpretation than these rules are edited by the designers of the system. This kind of system, even when its **legally simplistic**, led to a big improvement in **efficiency** and **quality** because the computer is perfectly suited for two difficulties that humans have. First, many legal rules have complex structures, the computer can calculate this perfectly. Second, human decision

makers were often not aware of relevant rules, the computer has a perfect memory. So, even when these systems are legally simplistic, they lead to a large increase of efficiency.

Casestudies Marlies van Eck
Does this mean that rule-based systems are the solution in administrative law? Not quite, because they still have some problems. This was identified by Marlies van Eck. She investigated two case studies, one for determining the right to child support and the other one determining fiscal income. She found several problems in the practice of these systems. The first one is that both systems were hard to explain, they were **untransparent**. Systems based on a symbolic approach are in general transparent and can be easily explained. The problem here is that the designers in practice have not followed the recommendations of the research community on how to make the systems explainable and transparent. So, lack of transparency is not only a problem for machine-learning systems but also for knowledge-based systems, at least if they are poorly designed. The research field of AI & Law has many recommendations for making a system explainable and transparent, but the engineers don't follow them. Second, there was **poor or missing documentation**. It was only readable for experts or not readable at all. Another problem was that in those cases were there was still a need for **interpretation** there was no human interference because the computer system was completely automatic. So, these interpretation problems had to be resolved in advance by the designers. Example of an ad-hoc interpretation of a legal rule: If application is more than seven days late, then inadmissible, unless there is a reasonable excuse. Implementation of 'reasonable excuse': If less then fifteen days late then reasonable excuse, otherwise not. Such interpretations should not be made by system designers but by legal decision makers. Another problem is that there are always **non- standard cases that do not fit the system.** A rule can be over or underinclusive. A final problem is that these systems are usually not stand-alone systems, but they are part of a chain of systems. If errors are made, then it is very hard to do this trough the chain.

Automating judicial routine decisions
So far, we have discussed applications of deductive rule-based expert systems in public administration. Can they also be used in court by judges? In 2017 the chairman of the Dutch Council of the Judiciary proposed that routine cases should be automated to give judges more time for non-trivial cases. He seemed to be thinking of outcome predictors (machine learning), but I have just told you that this is not a good idea. A better idea would be the rule-based decisions. But then the problem should not be to complex, it has to be small and well defined. Also there needs to be enough legal proof, no evidence problems. The facts can be automatically collected.

Limitations of legal rule-based systems
For the question whether the computer could solve hard legal cases these rule-based expert systems are clearly to limited. First of all, when there are serious issues about **proving the facts**, so it is not suitable for determining the facts from evidence. Second the system **cannot handle exceptions and rule conflicts**. There is also no way that they can account for the role of cases and precedents in legal decision making. In short there is no way an AI system can argue or make arguments like lawyers of judges do.

Part 3 Modelling legal reasoning – models of argumentation

AI & Law research on legal argument
What would a realistic AI-system do? The key is legal reasoning as **argumentation**. Inference by constructing and comparing arguments and counterarguments. And also leaving room for not just rule application but combining rules and precedents and appeal to principles and values. In this part of the lecture, I want to give you an introduction to the kind of models that have been presented.

Factor-based reasoning
Factor based problem domains are domains where there are **no clear rules**, but just collections of features or factors that are **tentative** reasons pro or con a conclusion. Often to different degrees. The job of the judge in such cases is to way the collections or factors that point **pro or con**, these decisions

becomes **precedents**. But how do judges **weigh** factors? And what if a new case does not perfectly **match** a precedent?

Example from US law: misuse of trade secrets
The HYPO-systems and CATO-systems are the most influential early systems. Both of these systems are for particular **factor-based problem domain** in US trade secret law. In particular the question whether there has been a misuse in the trade system. They went through the literature to identify the relevant factors pro and con. Some factors pro misuse of trade secrets: agreed not to disclose, security measures, identical products, knew info confidential. Some factors con misuse of trade secrets: disclosure in negotiations, info reverses engineerable.

HYPO-system
The output of the HYPO-system was a debate as a good common law lawyers could have. Where the first the **plaintiff** gets the turn, then the **defendant** and then the **plaintiff** again. The plaintiff can cite a precedent and share some relevant similarities with the case. Then the turn shifts to the defendant, he can do two things. First, the defendant can distinguish the precedent cited by the plaintiff, for instance by arguing that the case misses some pro plaintiff factors in the precedent, unlike in the present case etc. Another way to distinguish by pointing a new pro defendant factor in the new case that were not in the precedent. In addition to distinguishing the precedent the defendant can move a counter example, citing another precedent. The turn shifts to the plaintiff again, who can then try to distinguish all precedents cited by the defendant. The idea of the HYPO-system is that it would be a sparring partner for a solicitor so he could try out argumentation against the computer. The judge choses those precedents that best fits the case. The question arises if this differences and similarities at the factor-based level is that all there is or is there more to it? In this case, how can we decide cases? The idea of subsequence research is one idea that people can refer to underlying social or moral principles or values why one set of factors will outweigh another set off factors.

Ownership of wild animals
Especially the early research in AI and law on the use of values to decide cases uses a set of cases that are well known in law school, often used to teach law student how to argue in cases. All these cases are about the issue of ownership of wild animals. We will use three of them. The first one is **Pierson v. Post** where the plaintiff is **hunting a fox** on open land. The defendant kills the fox. The second one **Keeble v. Hickersgill** where the plaintiff is a **professional hunter**. He lured duck to his pond and the defendant scares the duck away with a shot in the air. The final case **Young v. Hitchens** where the plaintiff is a **professional fisherman**. He spread his nets and the defendant gets inside the larger nets and catches the fish. The issue in these cases is on which property the occasion occurred. The issue is how can we decide Young v Hitchens on the cases on these two precedents. First, we will analyze **on the base of the HYPO manner with factors** and then we will show how you can use social **values** to resolve this issue. This remind you on *Dworkins* idea what look to be hard cases can in the end be decided by underlying values.

Factors in the wild animal cases
First, we will give a HYPO style analyses by identifying the factors. Factors ate abstract patterns of relevant legal facts.

Pierson – won by defendant
- Defendant not pursuing livelihood (p1)
- Plaintiff not pursuing livelihood (d1)
- Plaintiff not on own land (d2)
- Plaintiff had not caught animal (d3)

Keeble – won by plaintiff
- Defendant not pursuing livelihood (p1)
- Plaintiff pursuing livelihood (p2)
- Plaintiff on own land (p3)
- Plaintiff had not caught animal (d3)

Young – won by defendant
- Defendant pursuing livelihood (d4)
- Plaintiff pursuing livelihood (p2)
- Plaintiff had not caught animal (d3)

Arguing Young in HYPO
So, what dispute would HYPO generate? The defendant could draw an analogy between young and piersson: 'I should win because as in Pierson, which was won by defendant, plaintiff was not hunting on his own land and had not caught the animal.' The plaintiff could **distinguish** by saying my case is different since unlike the plaintiff in *Pierson*, I was pursuing my livelihood as fisherman. The plaintiff could point at a **counterexample** that was won by the plaintiff. Plaintiff: 'I should win since my case is similar to Keeble, which was won by plaintiff because he was pursuing his livelihood even though he had not caught the animal.' The defendant could react by distinguishing Keeble by pointing at some relevant differences. 'Unlike defendant in Keeble, I am also pursuing my livelihood, and unlike in Keeble plaintiff is not hunting on his own land.' This is the HYPO style analyses. It's still a bit unfinished, we have to precedents which have some similarities and some differences, this is where the values come in.

Values promoted in the wild animal cases
Pierson – won by defendant
- Defendant not pursuing livelihood (p1)
- Plaintiff not pursuing livelihood (d1)
- Plaintiff not on own land (d2)
- Plaintiff had not caught animal (d3) promotes **certainty**

Keeble – won by plaintiff
- Defendant not pursuing livelihood (p1)
- Plaintiff pursuing livelihood (p2) promotes **economy**
- Plaintiff on own land (p3)
- Plaintiff had not caught animal (d3) promotes **certainty**

Young – (won by defendant)
- Defendant pursuing livelihood (d4) promotes **economy**
- Plaintiff pursuing livelihood (p2) promotes **economy**
- Plaintiff not on own land (d2) promotes **property**
- Plaintiff had not caught animal (d3) promotes **certainty**

Why is the factor that the plaintiff did not have yet caught the animal, why is that pro defendant factor? Because this **promotes certainty**, in the decision of Piersson if the outcome was who first saw the animal, this would lead to endless quarrel and litigation, and that is not good. We want to have a precise criterion. That is why the criteria is that the plaintiff had not caught the animal. This promotes certainty.

Why is the fact that the plaintiff was pursuing his livelihood a reason to decide for the plaintiff or defendant? That **promotes** the value of stimulating the **economy**. That is why this is pro plaintiff factor or pro defendant factor. So, this is a second social value promoting economic prosperity.

The final value has to do why the fact that the plaintiff was hunting on his own land as in Keeble, is a pro plaintiff factor, that is that it somehow **promotes** the sanctity of **property**.

How can we explain the outcome of the three cases in terms of the underlying values? In the Piersson case we see that deciding for the plaintiff would not promote any value, while deciding for the defendant promotes the value of legal certainty. So, this explains the outcome in this case, the defendant won because that promotes a value. In the Keeble case it is more subtle. The deciding for the plaintiff promotes both the value of economic property and sanctity of property while the defendant only

promotes the legal certainty. We can say that the court found the combination of promoting the economy and property outweighing the value of certainty. Young was won by the defendant because the plaintiff would only promote the value of economy and the defendant promotes economy and certainty. We don't count but the fact that deciding for the defendant promotes all values promoted by deciding for the plaintiff and some more values as well. This shows how judges often appeal to underlying values in order to justify decisions.

Legal argumentation systems: the KA bottleneck (KA = knowledge acquisition)
So far, these models are hardly begin applied in practice, because of the required knowledge is hard to manually acquire and code. You could say the computer could learn itself from the case law and law journals.

Oefenvragen
1. Men zegt wel eens: "wat er ook zij van de problemen met algoritmische uitkomstvoorspellers, zodra ze bijna 100% accuraat zijn verdwijnen deze problemen en kunnen we op hen vertrouwen". Wat vind je van deze stelling?
2. Welke soorten kennis zou een AI-systeem nodig hebben om in 'hard cases' te beslissen? (en dan volgende week: in welke mate is het mogelijk voor de computer om deze kennis automatisch te leren uit data?)
3. Kun je een factorgebaseerd probleemgebied noemen uit het Nederlandse recht?