logic and explainable ai in the legal domain

luís cruz-filipe

(joint work with jonas vistrup et al.)

department of mathematics and computer science
university of southern denmark

seminário de lógica matemática
january 6th, 2023

the goal

long-term

build an expert system for a (subdomain) of danish law

- current target: traffic law

the goal

long-term

build an expert system for a (subdomain) of danish law

- current target: traffic law

in this talk

overview of the project and initial steps

- motivation and relevance
- state of the art
- work plan and expected challenges

the goal

long-term

build an expert system for a (subdomain) of danish law

- current target: traffic law

in this talk

overview of the project and initial steps

- motivation and relevance
- state of the art
- work plan and expected challenges

and also

some more general considerations about logic and ai

ai throughout the years

general goal

build an "intelligent agent"

ai throughout the years

general goal

build an "intelligent agent"

- in hindsight, very, very, VERY naive
- behaviour is perceived as rational
- turing test

ai throughout the years

general goal

build an "intelligent agent"

- in hindsight, very, very, VERY naive
- behaviour is perceived as rational
- turing test

limitation

"success" is defined via observable behaviour

- relatively easy to fake
- does not require a model of the world

some disappointing success stories

eliza

- developed in the 1960s to show the superficiality of human/machine communication
- effectively the first chatbot
- no memory, but perceived as very human...

some disappointing success stories

eliza

- developed in the 1960s to show the superficiality of human/machine communication
- effectively the first chatbot
- no memory, but perceived as very human...

deep blue

- chess-playing program developed from 1985 to 1997
- first program able to beat a human champion
- ...but disappointing in many ways

machine learning

the trendy ai

machine learning is the major field of research within ai
for (too) many it is synonymous with ai

- very impressive results
- wide applicability: automatic translation, natural language processing, computer vision
- current hype: openai chat program

machine learning

the trendy ai

machine learning is the major field of research within ai
for (too) many it is synonymous with ai

- very impressive results
- wide applicability: automatic translation, natural language processing, computer vision
- current hype: openai chat program

basic idea

develop programs that can learn from data

- deductive methods, e.g. in sat solving
- statistical methods, e.g. deep learning

machine learning, cont'd

the dark side

machine learning is not the universal panacea

- no semantics!
- can be tricked, reproduces bias
- good at mimicking, but what is really going on?

machine learning, cont'd

the dark side

machine learning is not the universal panacea

- no semantics!
- can be tricked, reproduces bias
- good at mimicking, but what is really going on?

alternatives

go back to the basics!

- logic-based techniques
- explicit, understandable models of the real world

$\rightsquigarrow$ less effective, more trustworthy

new requirements for ai

trustworthy ai

using black-box systems can pose ethical issues

- medical diagnosis or treatment
- decision-making in with social consequences
- the legal domain

new requirements for ai

trustworthy ai

using black-box systems can pose ethical issues

- medical diagnosis or treatment
- decision-making in with social consequences
- the legal domain

the push for explainability

increasing requirements for systems that can:

- explain their decisions
- argue for them/defease arguments against them
- be inspected and questioned
- revise their beliefs

the challenge of ai in law

hard constraints

ethical issues are central in the legal domain, but there are others:

- tradition for argumentation
- counselling is accepted, ordering is not

the challenge of ai in law

hard constraints

ethical issues are central in the legal domain, but there are others:

- tradition for argumentation
- counselling is accepted, ordering is not

the current reality

law is nearly absent from the realm of applications of ai

- this project wants to change that!

a recent field of interest

logical approaches

logics for modeling law – modal logics and argumentation frameworks, among others

$\rightsquigarrow$ mostly descriptive systems

a recent field of interest

logical approaches

logics for modeling law – modal logics and argumentation frameworks, among others

$\rightsquigarrow$ mostly descriptive systems

formalization approaches

the french tax system has been formalized in coq

- an interesting and large-scale application
- on the borderline of what is considered ai

a recent field of interest

logical approaches

logics for modeling law – modal logics and argumentation frameworks, among others

$\rightsquigarrow$ mostly descriptive systems

formalization approaches

the french tax system has been formalized in coq

- an interesting and large-scale application
- on the borderline of what is considered ai

the bad examples (usa)

machine-learning systems for counselling judges on sentencing

our project

the goal

build a logic-based expert system for a (subdomain of) danish law

our project

the goal

build a logic-based expert system for a (subdomain of) danish law

current toy example

simple datalog solver based on sld-resolution, implemented in java

- clauses have a textual translation
- traces of sld-resolution can be presented in natural language and inspected on-demand
- two (very unrealistic) rules, a set of facts
- proof-of-concept: the lawyers' reaction

our project

the goal

build a logic-based expert system for a (subdomain of) danish law

target prototype

full formalization of danish traffic law

- 60 pages of law
- 1200 pages of explanation...
- mostly deterministic, consistent and self-contained
- but: some instances of defaults and exceptions

expected challenges

law is sometimes illogical

need for more complex forms of reasoning

- default reasoning
- paraconsistent reasoning
- hypothetical and abductive reasoning

expected challenges

law is sometimes illogical

need for more complex forms of reasoning

- default reasoning
- paraconsistent reasoning
- hypothetical and abductive reasoning

tools and techniques

logic programming is a good starting point

- it has default negation
- rule-based models for default reasoning
- possibility of defining negation as separate predicate
- work on abduction and hypothetical reasoning

additional goals

from our legal friends

- evaluate the system in practice
- argue for the feasibility of using ai in the legal domain

from our social sciences friends

- understand the interplay beween humans and machines
- study the mutual effects of ai systems and humans in each others' behaviours

thank you!