Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

"Data! Data! Data!"

Sherlock Holmes

AI in forensics

Rolf Ypma

Principal Scientist | Forensic Data Scientist

Netherlands Forensic Institute

r.ypma at nfi.nl

Uses of AI (algorithms)

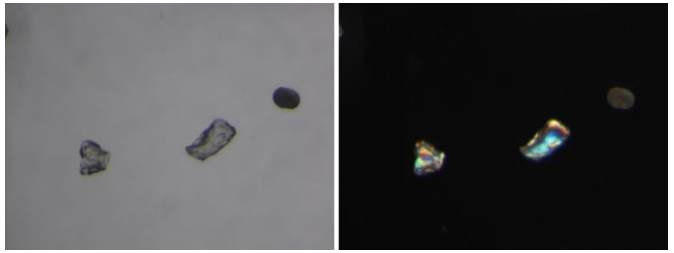
- Supportive processes / automation
  - Mostly internal/low impact
- Finding traces
- Evidence evaluation

Fast automated results

- DNA (and drugs)
- Initial identification effort in three days, rather than two weeks
  - No machine learning
  - Well understood algorithms
  - A lot of testing/validation
  - Slow stepwise introduction
- Described in the ministry of Justice ‘algorithm registry’

Finding traces

Examples	Score
Mate he wil visit and luere him ans they will **shoot** him!	0.94
He has to **disapppear,** it is enough	0.99
No worries mate he will **sleep** tomorrow	0.99
Oh **shoot** I forgot that	0.00
Your friends **sleep** over?	0.00



Next: physical traces

Finding traces - considerations

- Model
  - Development (or procurement)
    - Base model provenance/datasets
    - Data
  - Validation
    - When is it fit for use?
    - Data
  - Deployment
    - Who will use it how for what?
    - Training/documentation/model cards

Finding traces - considerations

- What could go wrong?
  - Traces missed. Exhonerating traces?
  - Overreliance on computer. Criminal message because computer says so?
  - Bias. Messages predominantly found for one demographic as model was trained on these?

Evidence evaluation

Likelihood Ratio (LR) = P(E|H1) / P(E|H2)

‘Does the evidence fit the first or the second hypothesis better?’

Evidence evaluation - DNA

Likelihood Ratio (LR) = P(E|H1) / P(E|H2)

‘Does the evidence fit the first or the second hypothesis better?’

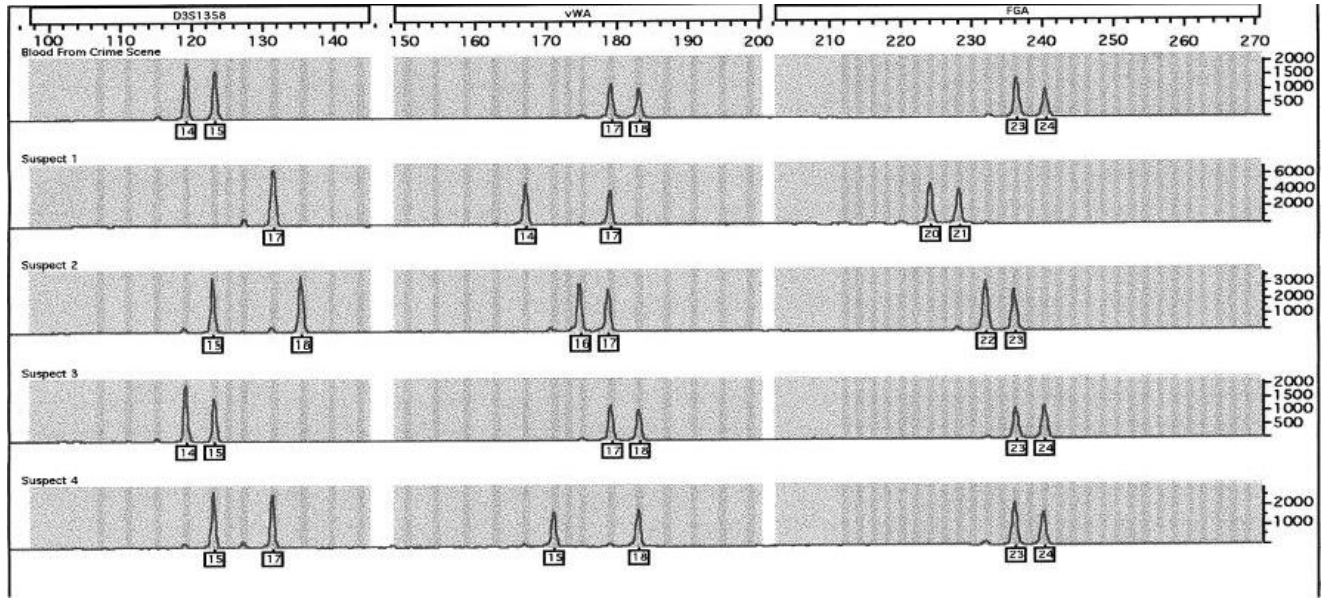
E: the measurements on the DNA

H1: suspect is the donor of the DNA

H2: someone else is the donor of the DNA

Evidence evaluation

- Algorithmic evaluation has been a goal for two decades
- To alleviate bias



Thompson (2009)

Evidence evaluation – DNA

Likelihood Ratio (LR) = P(E|H1) / P(E|H2)

~ 1 / frequency of this DNA profile

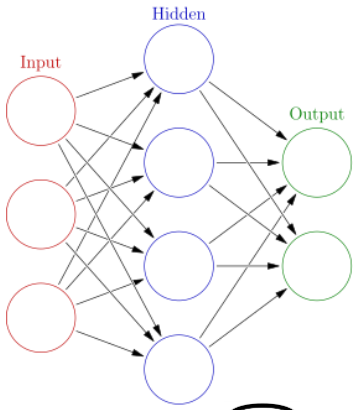
Few and interpretable parameters

Relatively predictable output

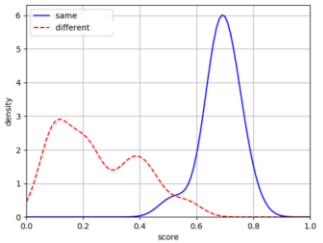
'Explainable'

AI (eg deep learning) in evidence evaluation

validation



0.3



1/100

report

questioned pair

AI
'black box'

score

calibration

LR

expert

Fit for purpose?

- Verdicts can be based to a large extent on algorithm output
- Are AI algorithms reliable (enough)?
- Opinion: explainability is not the answer

Evidence evaluation – biometry

No ‘simple’ profile, but audio/image files

‘Simple’ algorithms perform poorly

Deep learning (‘AI’) performs well
= ‘Ridiculously large mathematical equations, based on data rather than knowledge’

What does this parameter do? What will the output be for this image?

- we will not know until when we test it
- so we test it

Analogy

	Forensic DNA analysis	Medicinal drugs
Users	Lawyers/triers of fact	Patients
Practitioners	Forensic DNA experts	Doctors
Developers	Statisticians	Bioengineers
Trust	Understanding	Clinical trials

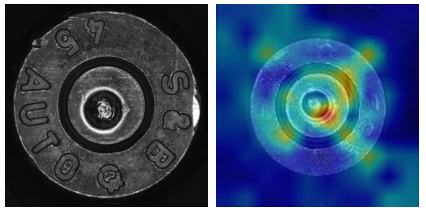
Conclusion

- AI/algorithms are used throughout forensics
  - Impact in automated output / evidence evaluation

- We think AI can be used responsibly in evidence evaluation
  - Strong validation
  - Strong human oversight
  - Explainability where possible

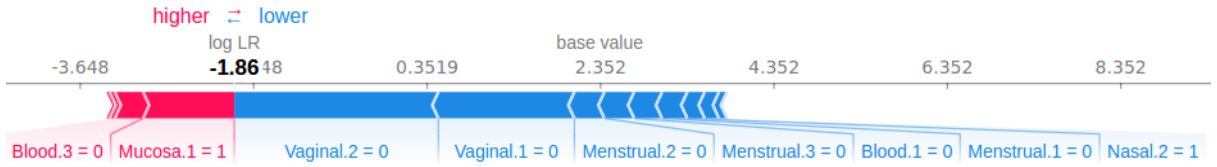
Explainability

- Uses
  - Increase trust
  - Enable improvements
  - Rule out applicability



Cartridge cases

- Notes
  - Phrenology and lie detectors were ‘explained’
  - Experts are not



mRNA

Ypma et al. (2021) FSI Gen