Evaluating Legal Accuracy of Neural Generators on the Generation of Criminal Court Dockets Description

Nicolas Garneau, Eve Gaumond,
Luc Lamontagne and Pierre-Luc Déziel

INLG 2022, July 19th

Team

Nicolas

Eve

Luc

Pierre-Luc

iid
Institut intelligence et données

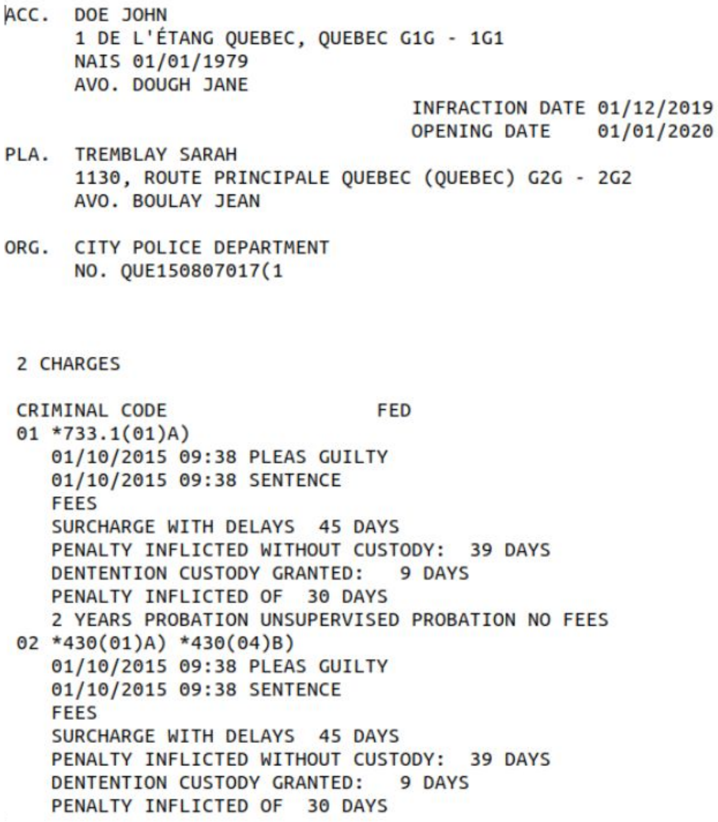
OBSERVATOIRE INTERNATIONAL SUR LES IMPACTS SOCIÉTAUX DE L'IA ET DU NUMÉRIQUE

*SOQUIJ
Intelligence juridique

CRSH SSHRC

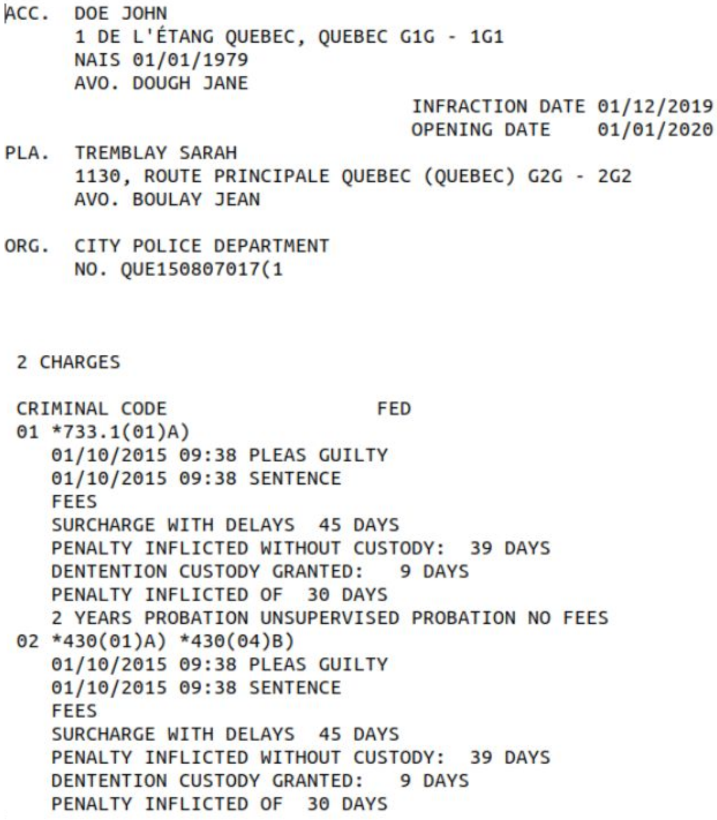
NSERC CRSNG

Court Dockets



```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE
                                  INFRACTION DATE 01/12/2019
                                  OPENING DATE    01/01/2020
PLA.   TREMBLAY SARAH
       1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
       AVO. BOULAY JEAN

ORG.   CITY POLICE DEPARTMENT
       NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

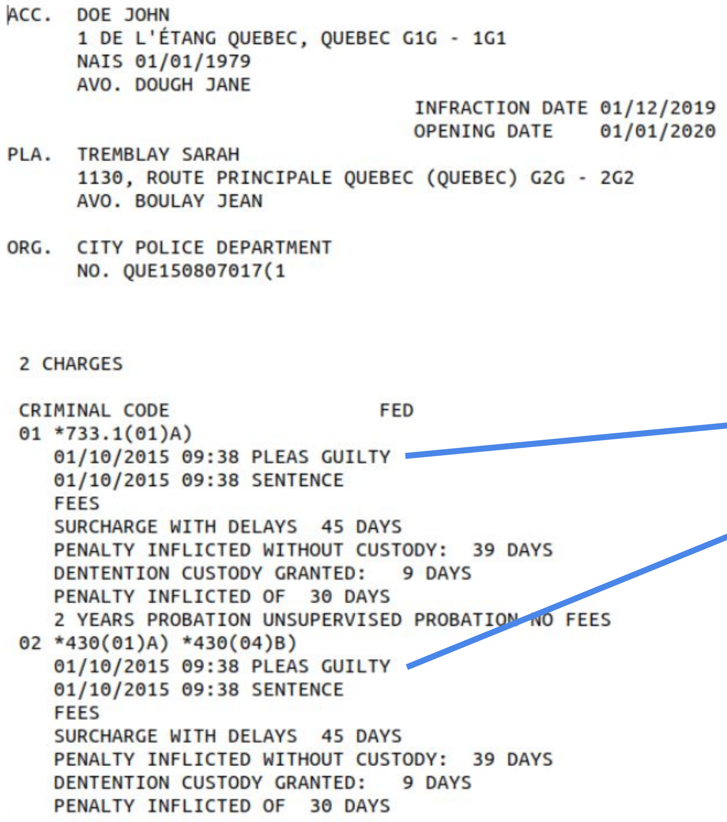
Court Dockets



```
ACC.  DOE JOHN
      1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
      NAIS 01/01/1979
      AVO. DOUGH JANE
                             INFRACTION DATE 01/12/2019
                             OPENING DATE    01/01/2020
PLA.  TREMBLAY SARAH
      1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
      AVO. BOULAY JEAN

ORG.  CITY POLICE DEPARTMENT
      NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

"John Doe ***pleaded guilty*** *to* ***failure to comply with an order*** *and* ***mischief to property*** *on* **October 1st, 2015***."*

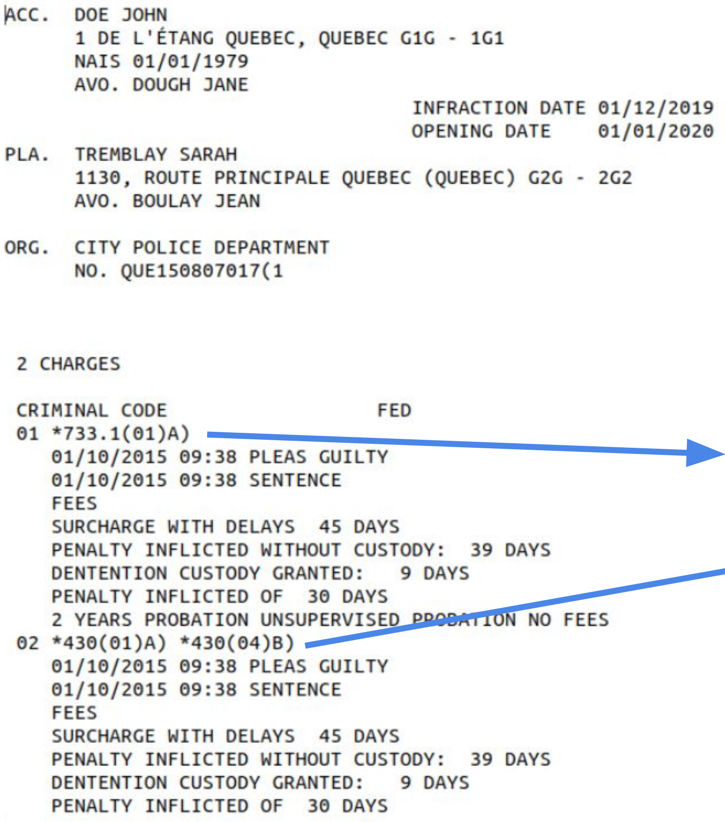
Court Dockets



```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE
                                 INFRACTION DATE 01/12/2019
                                 OPENING DATE    01/01/2020
PLA.   TREMBLAY SARAH
       1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
       AVO. BOULAY JEAN

ORG.   CITY POLICE DEPARTMENT
       NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

"John Doe ***pleaded guilty*** *to* ***failure to comply with an order*** *and* ***mischief to property*** *on* ***October 1st, 2015****."*

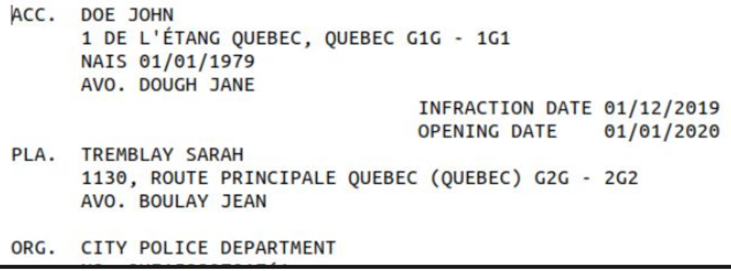
Court Dockets



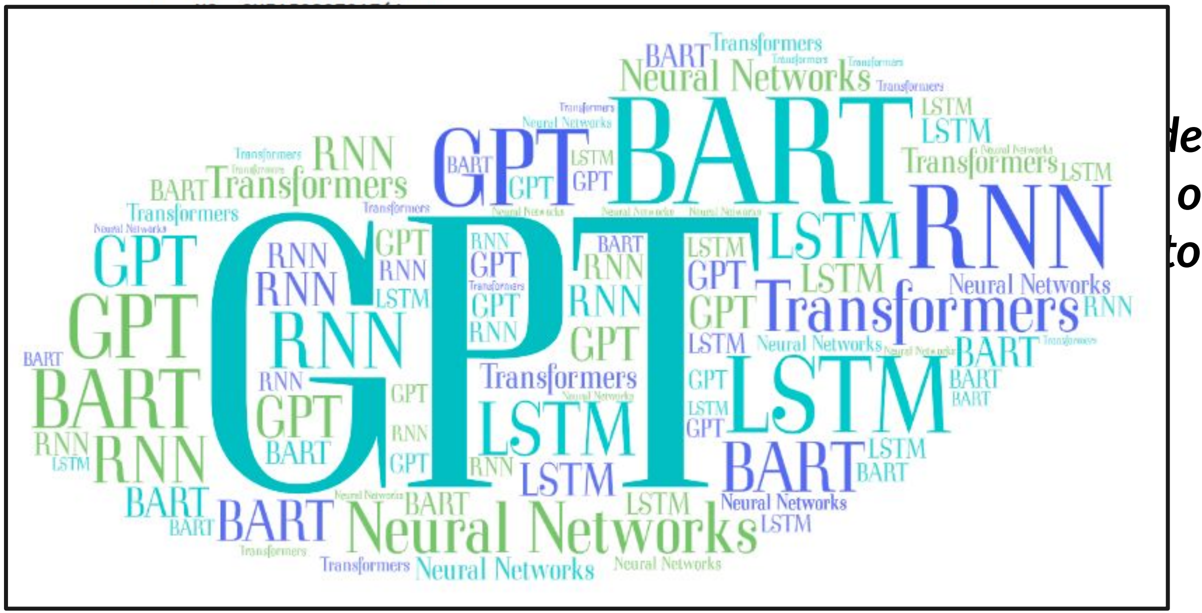
```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE
                                 INFRACTION DATE 01/12/2019
                                 OPENING DATE    01/01/2020
PLA.   TREMBLAY SARAH
       1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
       AVO. BOULAY JEAN

ORG.   CITY POLICE DEPARTMENT
       NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

“John Doe ***pleaded guilty*** *to* ***failure to comply with an order*** *and* ***mischief to property*** *on* ***October 1st, 2015****.”*

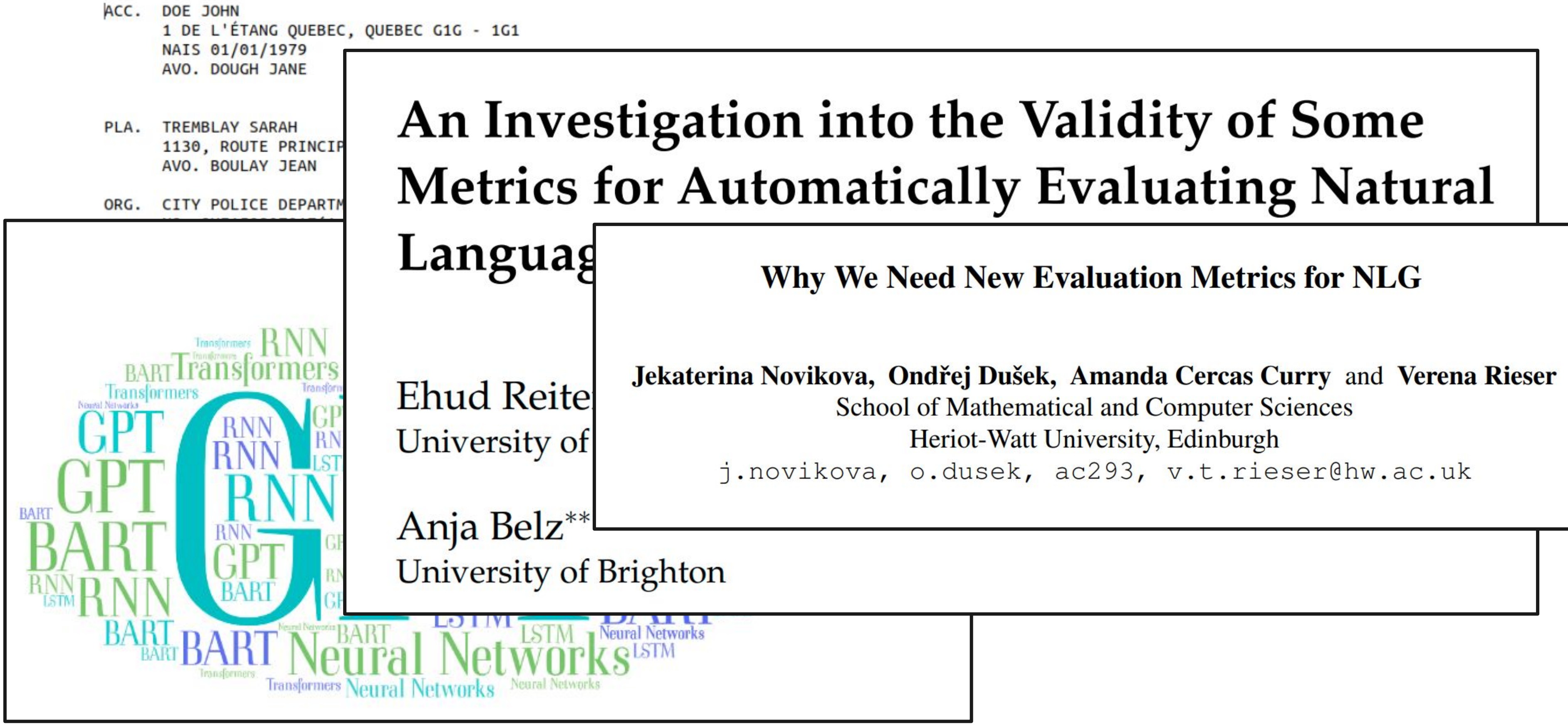
Court Dockets



```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE
                                   INFRACTION DATE 01/12/2019
                                   OPENING DATE    01/01/2020
PLA.   TREMBLAY SARAH
       1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
       AVO. BOULAY JEAN

ORG.   CITY POLICE DEPARTMENT
```



led guilty *to* ***failure to***
order *and* ***mischief to***
tober 1st, 2015*.”*



```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE

PLA.   TREMBLAY SARAH
       1130, ROUTE PRINCIP
       AVO. BOULAY JEAN

ORG.   CITY POLICE DEPARTM
```

Transformers
RNN
BART
GPT
LSTM
Neural Networks
G

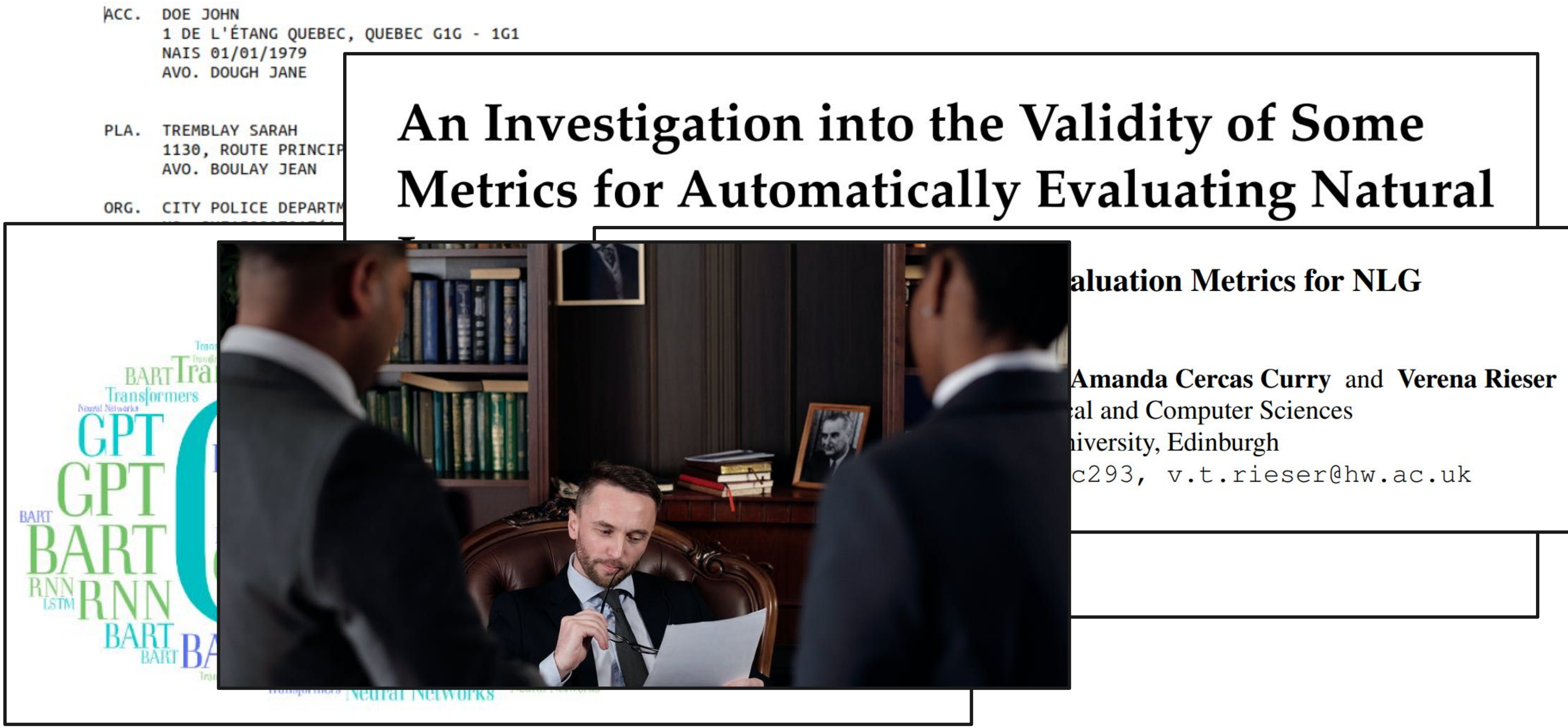
An Investigation into the Validity of Some Metrics for Automatically Evaluating Natural Languag

Ehud Reite
University of

Anja Belz**
University of Brighton

Why We Need New Evaluation Metrics for NLG

Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry and **Verena Rieser**
School of Mathematical and Computer Sciences
Heriot-Watt University, Edinburgh
j.novikova, o.dusek, ac293, v.t.rieser@hw.ac.uk



```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE

PLA.   TREMBLAY SARAH
       1130, ROUTE PRINCIP
       AVO. BOULAY JEAN

ORG.   CITY POLICE DEPARTM
```

An Investigation into the Validity of Some Metrics for Automatically Evaluating Natural

aluation Metrics for NLG

Amanda Cercas Curry and Verena Rieser
al and Computer Sciences
iversity, Edinburgh
c293, v.t.rieser@hw.ac.uk

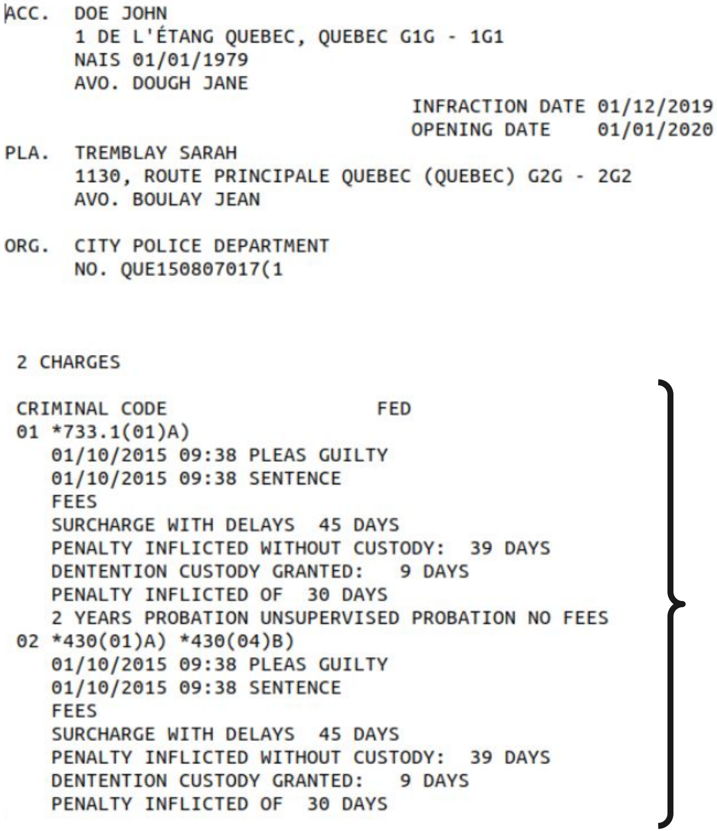
```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE

PLA.   TREMBLAY SARAH
       1130, ROUTE PRINCIP
       AVO. BOULAY JEAN

ORG.   CITY POLICE DEPARTM
```

An Investigation into the Validity of Some Metrics for Automatically Evaluating Natural

aluation Metrics for NLG

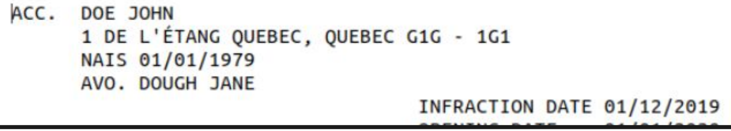
Amanda Cercas Curry and Verena Rieser
al and Computer Sciences
iversity, Edinburgh
c293, v.t.rieser@hw.ac.uk

Court Dockets

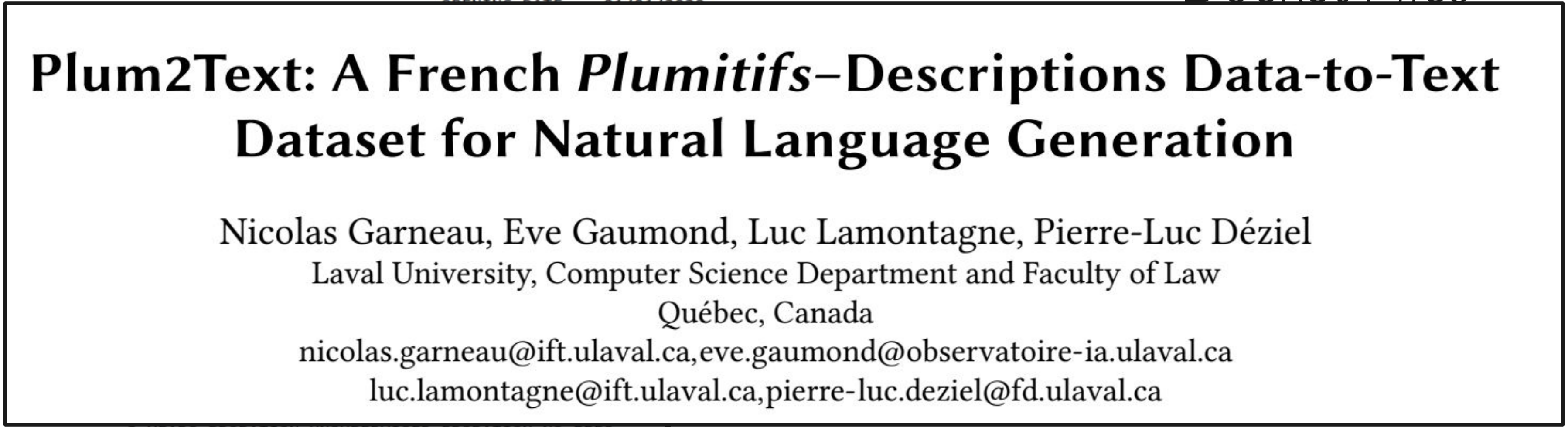


```
ACC.  DOE JOHN
      1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
      NAIS 01/01/1979
      AVO. DOUGH JANE
                             INFRACTION DATE 01/12/2019
                             OPENING DATE    01/01/2020
PLA.  TREMBLAY SARAH
      1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
      AVO. BOULAY JEAN

ORG.  CITY POLICE DEPARTMENT
      NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

Docket Files

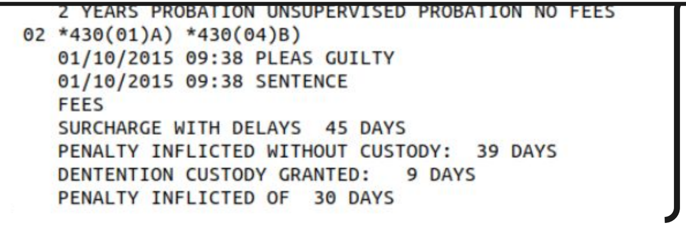


```
ACC.   DOE JOHN
       1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
       NAIS 01/01/1979
       AVO. DOUGH JANE
                                 INFRACTION DATE 01/12/2019
```



Plum2Text: A French *Plumitifs*–Descriptions Data-to-Text Dataset for Natural Language Generation

Nicolas Garneau, Eve Gaumond, Luc Lamontagne, Pierre-Luc Déziel
Laval University, Computer Science Department and Faculty of Law
Québec, Canada
nicolas.garneau@ift.ulaval.ca,eve.gaumond@observatoire-ia.ulaval.ca
luc.lamontagne@ift.ulaval.ca,pierre-luc.deziel@fd.ulaval.ca



```
     2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
  02 *430(01)A) *430(04)B)
     01/10/2015 09:38 PLEAS GUILTY
     01/10/2015 09:38 SENTENCE
     FEES
     SURCHARGE WITH DELAYS  45 DAYS
     PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
     DENTENTION CUSTODY GRANTED:   9 DAYS
     PENALTY INFLICTED OF  30 DAYS
```

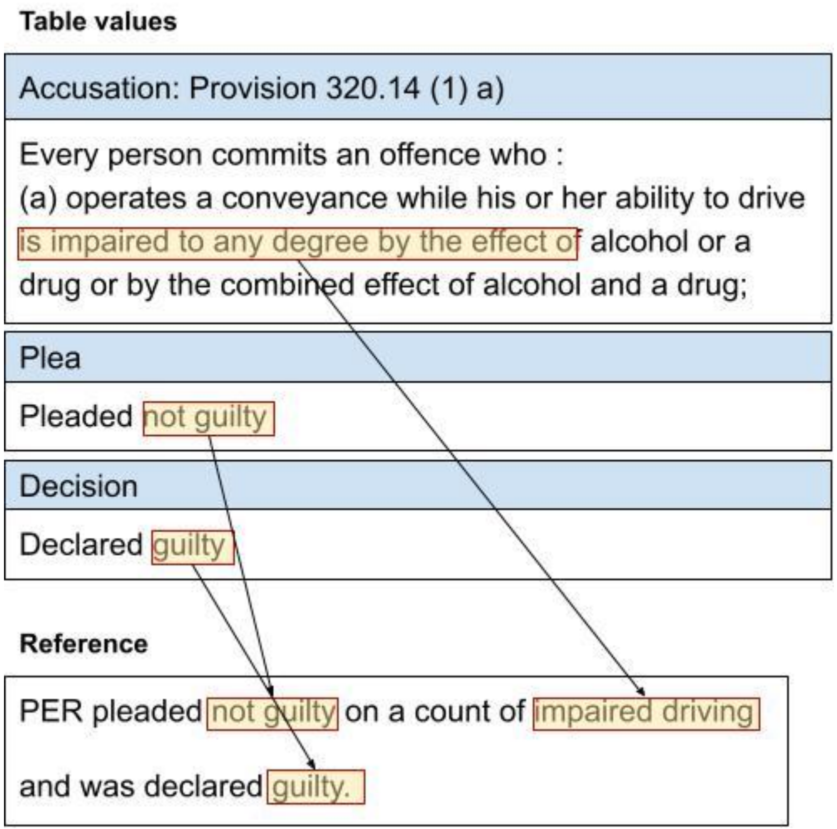


Table values

Accusation: Provision 320.14 (1) a)
Every person commits an offence who : (a) operates a conveyance while his or her ability to drive is impaired to any degree by the effect of alcohol or a drug or by the combined effect of alcohol and a drug;

Plea
Pleaded not guilty

Decision
Declared guilty

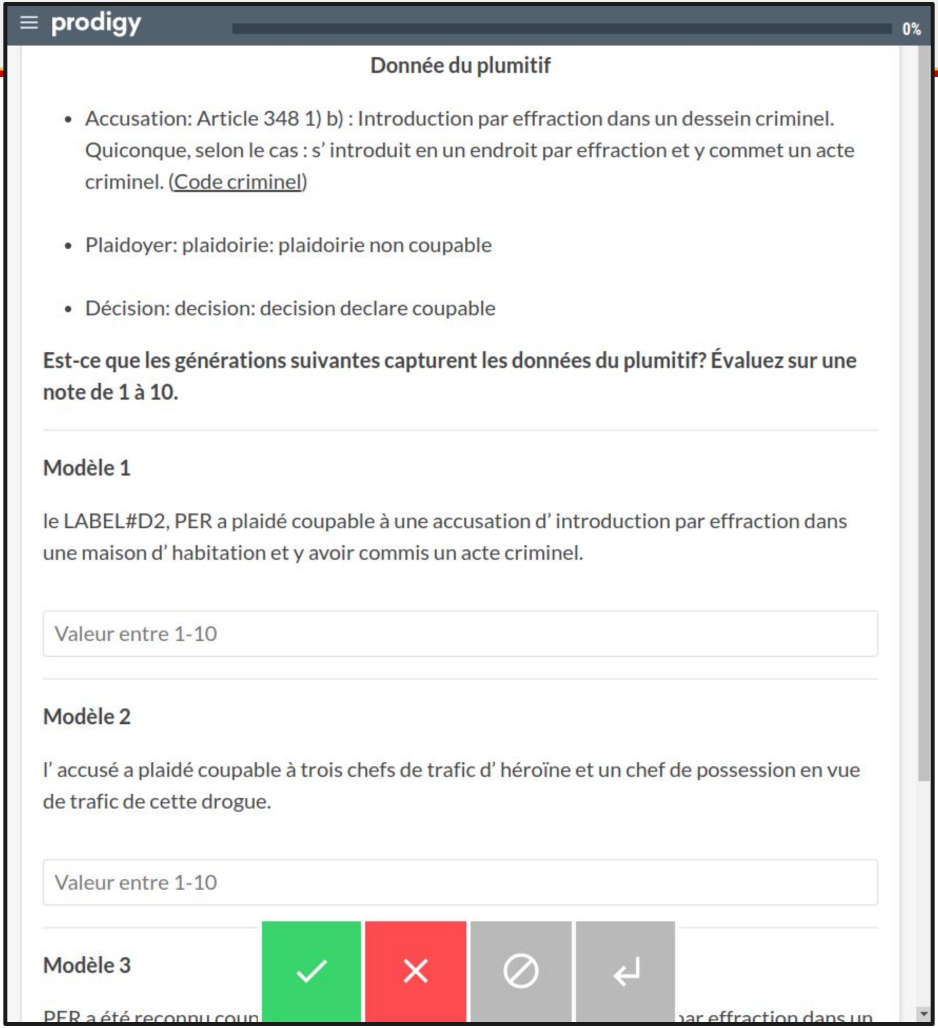
Reference

PER pleaded not guilty on a count of impaired driving

and was declared guilty.

We trained 3 models with different priors

1. LSTM from scratch (*no prior*)
2. BARThez (*language prior*)
3. *CriminelBART* (*language and domain prior*)



Donnée du plumitif

- Accusation: Article 348 1) b) : Introduction par effraction dans un dessein criminel. Quiconque, selon le cas : s’ introduit en un endroit par effraction et y commet un acte criminel. (Code criminel)
- Plaidoyer: plaidoirie: plaidoirie non coupable
- Décision: decision: decision declare coupable

Est-ce que les générations suivantes capturent les données du plumitif? Évaluez sur une note de 1 à 10.

Modèle 1

le LABEL#D2, PER a plaidé coupable à une accusation d’ introduction par effraction dans une maison d’ habitation et y avoir commis un acte criminel.

Valeur entre 1-10

Modèle 2

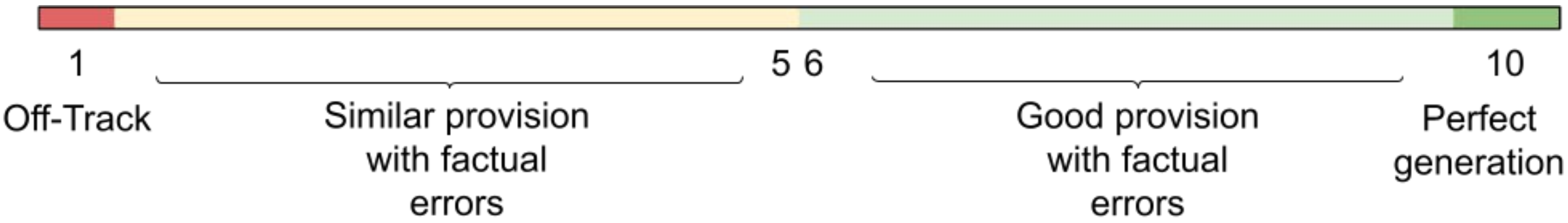
l’ accusé a plaidé coupable à trois chefs de trafic d’ héroïne et un chef de possession en vue de trafic de cette drogue.

Valeur entre 1-10

Modèle 3

PER a été reconnu coup par effraction dans un

Legal Accuracy Scale



1

Off-Track

Similar provision with factual errors

5 6

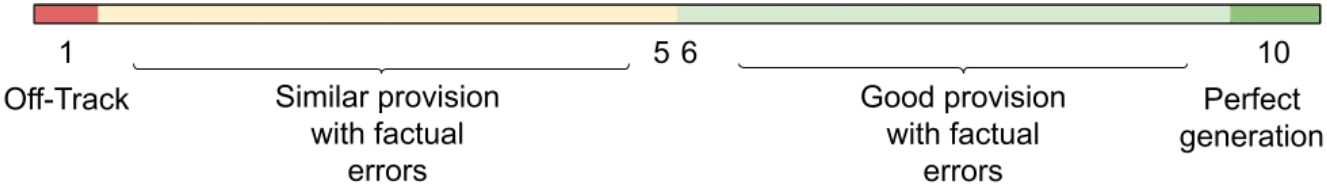
Good provision with factual errors

10

Perfect generation

Legal Accuracy Scale

1. **Theme**: some provisions are similar to others (e.g. *Trafficking in substance* and *Possession for purpose of trafficking*)
    - Position on the scale
2. **Precision (factual errors)**:
    - Hallucination: Anything not supported by the table
    - Omission: Table value not verbalized
        - Points on the scale


1
Off-Track
Similar provision with factual errors
5 6
Good provision with factual errors
10
Perfect generation

Results of the human evaluation

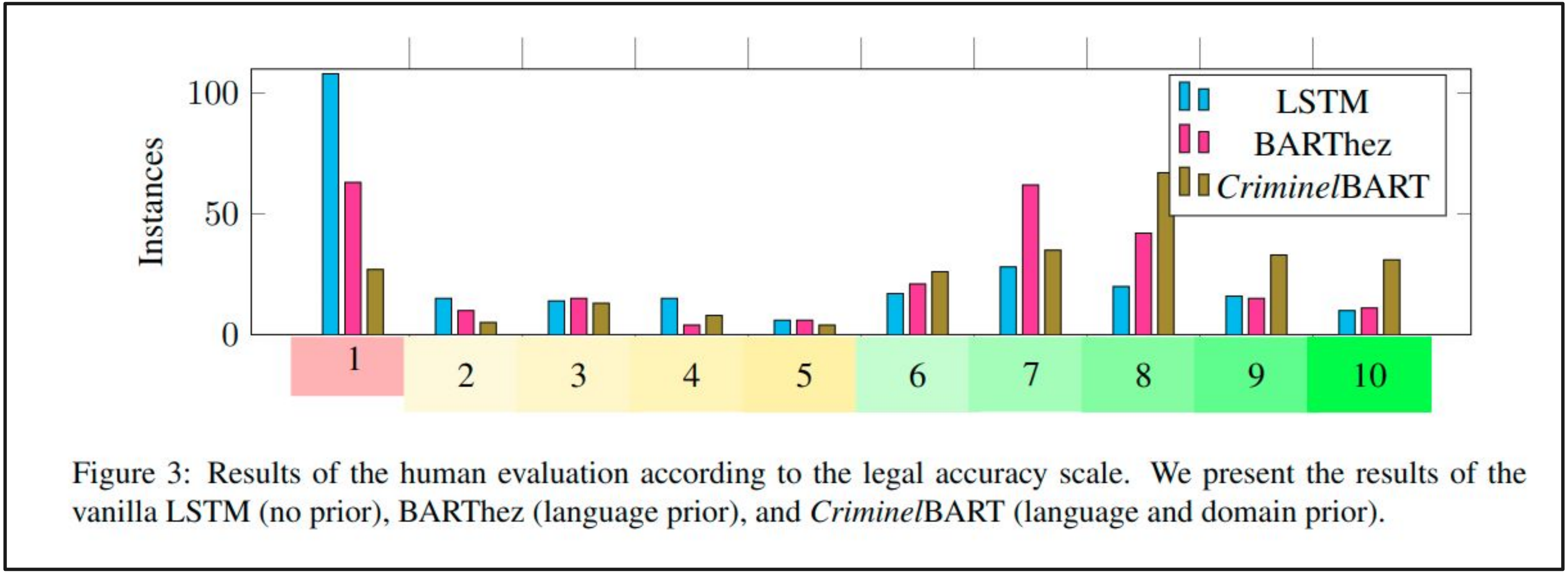


Figure 3: Results of the human evaluation according to the legal accuracy scale. We present the results of the vanilla LSTM (no prior), BARThez (language prior), and *Criminel*BART (language and domain prior).

Results of the human evaluation

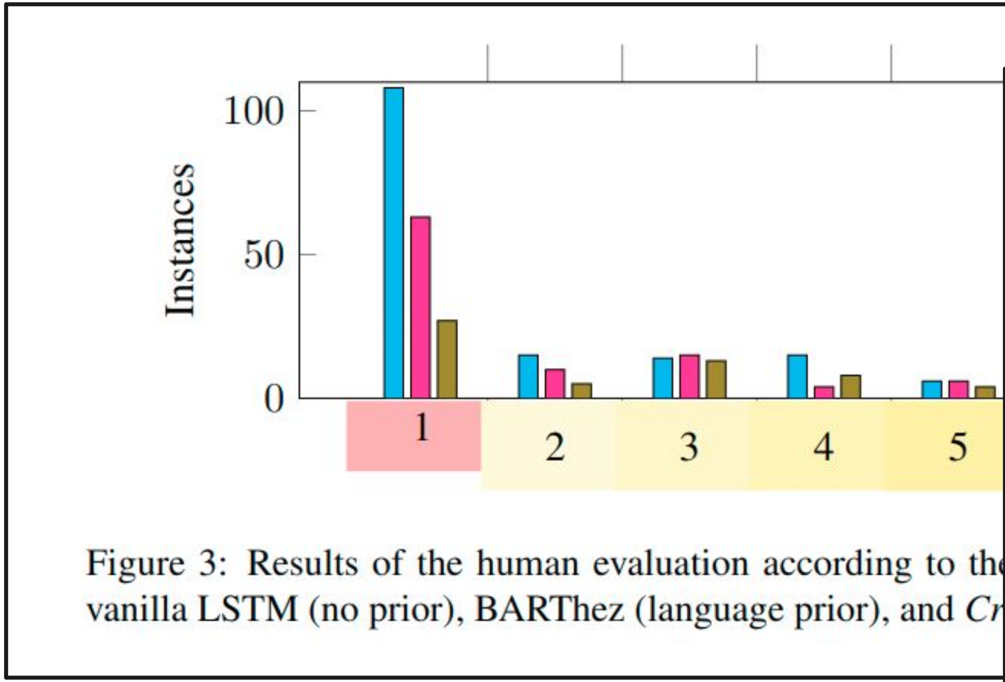
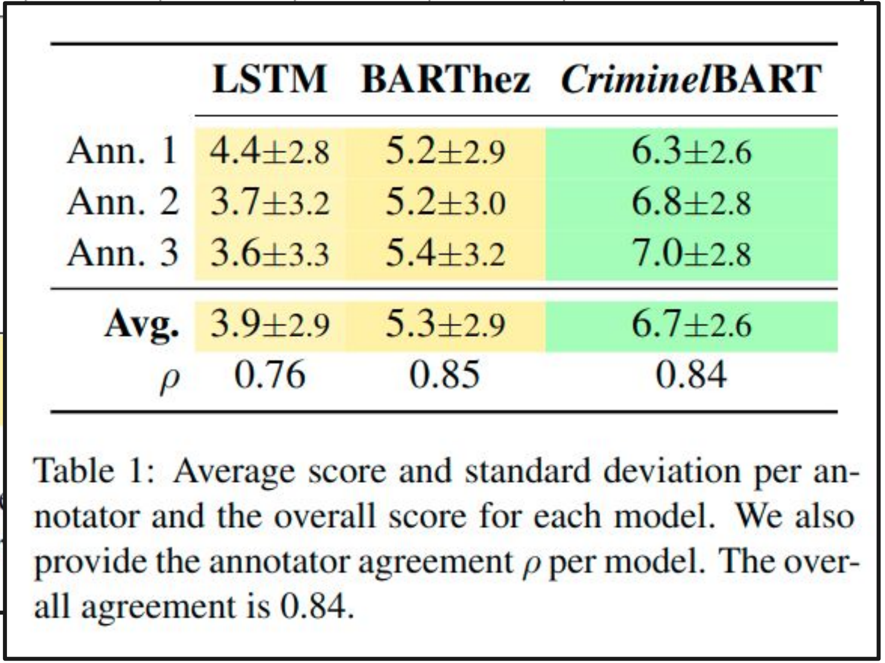


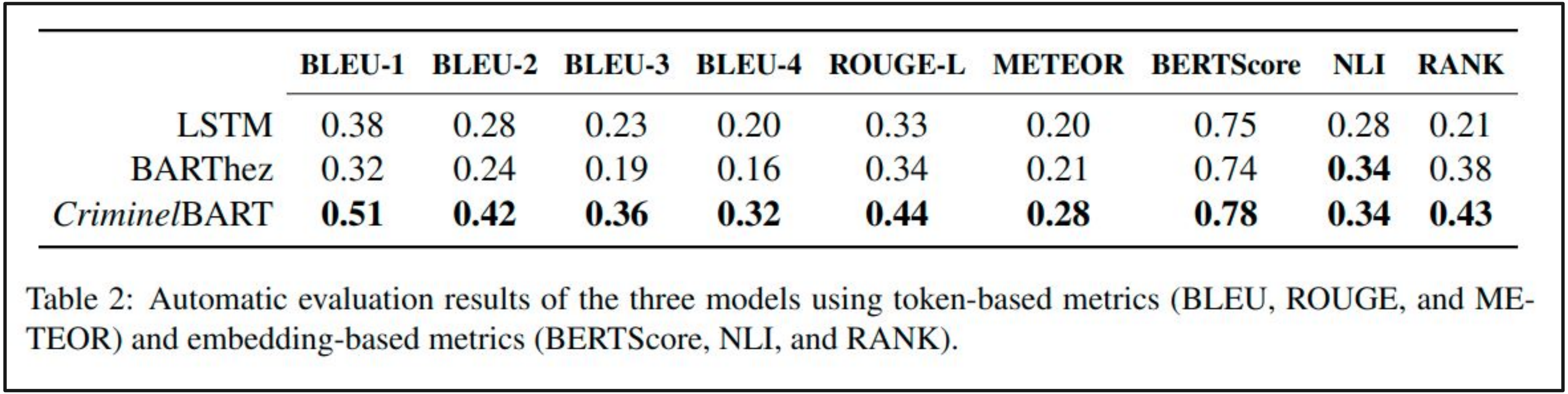
Figure 3: Results of the human evaluation according to the vanilla LSTM (no prior), BARThez (language prior), and Cr



	LSTM	BARThez	*Criminel*BART
Ann. 1	4.4±2.8	5.2±2.9	6.3±2.6
Ann. 2	3.7±3.2	5.2±3.0	6.8±2.8
Ann. 3	3.6±3.3	5.4±3.2	7.0±2.8
Avg.	3.9±2.9	5.3±2.9	6.7±2.6
ρ	0.76	0.85	0.84

Table 1: Average score and standard deviation per annotator and the overall score for each model. We also provide the annotator agreement ρ per model. The overall agreement is 0.84.

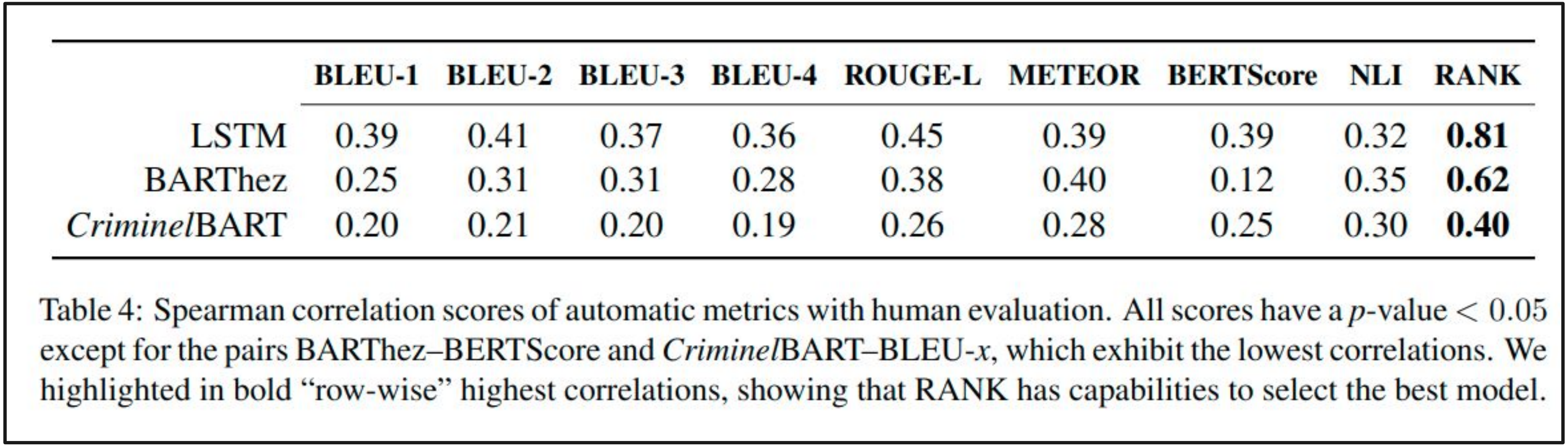
Automatic evaluation results



	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	METEOR	BERTScore	NLI	RANK
LSTM	0.38	0.28	0.23	0.20	0.33	0.20	0.75	0.28	0.21
BARThez	0.32	0.24	0.19	0.16	0.34	0.21	0.74	**0.34**	0.38
*Criminel*BART	**0.51**	**0.42**	**0.36**	**0.32**	**0.44**	**0.28**	**0.78**	**0.34**	**0.43**

Table 2: Automatic evaluation results of the three models using token-based metrics (BLEU, ROUGE, and METEOR) and embedding-based metrics (BERTScore, NLI, and RANK).

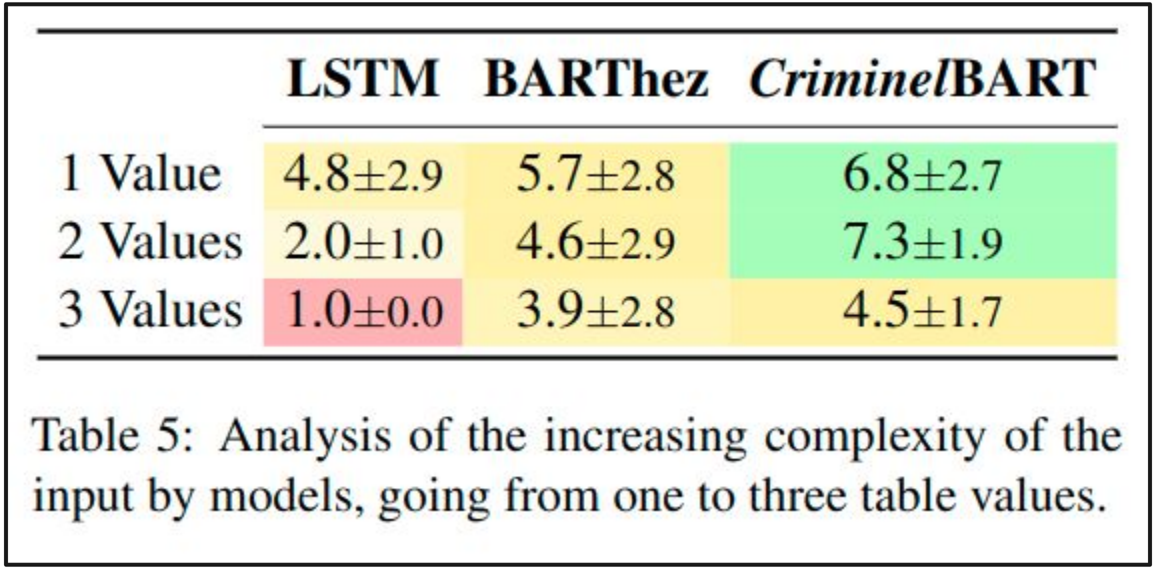
Correlation between automatic vs human evaluation



	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	METEOR	BERTScore	NLI	RANK
LSTM	0.39	0.41	0.37	0.36	0.45	0.39	0.39	0.32	**0.81**
BARThez	0.25	0.31	0.31	0.28	0.38	0.40	0.12	0.35	**0.62**
*Criminel*BART	0.20	0.21	0.20	0.19	0.26	0.28	0.25	0.30	**0.40**

Table 4: Spearman correlation scores of automatic metrics with human evaluation. All scores have a p-value < 0.05 except for the pairs BARThez–BERTScore and *Criminel*BART–BLEU-x, which exhibit the lowest correlations. We highlighted in bold "row-wise" highest correlations, showing that RANK has capabilities to select the best model.

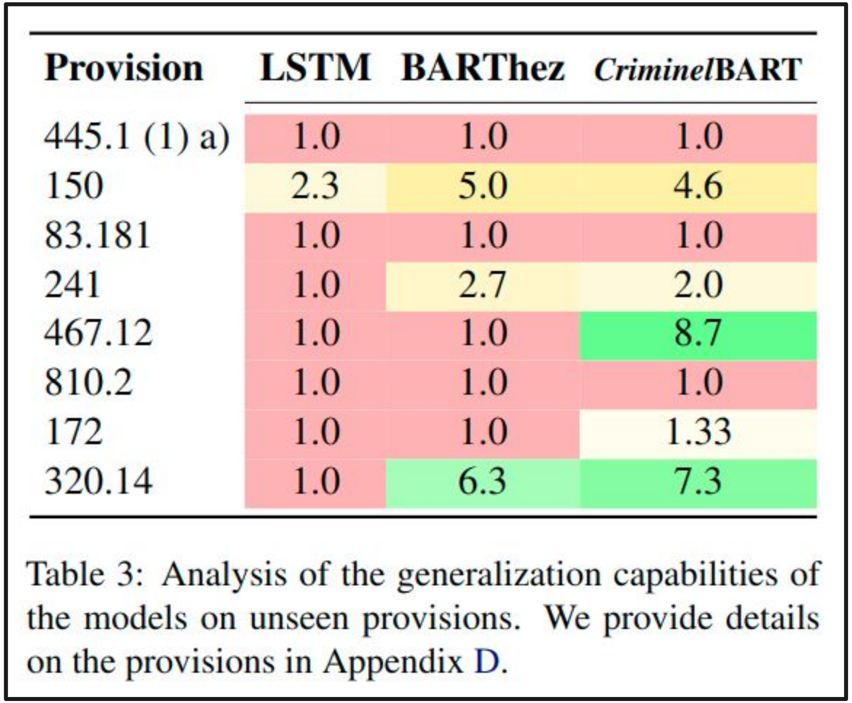
Increasing complexity of the input



	LSTM	**BARThez**	***Criminel*BART**
1 Value	4.8±2.9	5.7±2.8	6.8±2.7
2 Values	2.0±1.0	4.6±2.9	7.3±1.9
3 Values	1.0±0.0	3.9±2.8	4.5±1.7

Table 5: Analysis of the increasing complexity of the input by models, going from one to three table values.

Impact of prior knowledge



Provision	LSTM	BARThez	*Criminel*BART
445.1 (1) a)	1.0	1.0	1.0
150	2.3	5.0	4.6
83.181	1.0	1.0	1.0
241	1.0	2.7	2.0
467.12	1.0	1.0	8.7
810.2	1.0	1.0	1.0
172	1.0	1.0	1.33
320.14	1.0	6.3	7.3

Table 3: Analysis of the generalization capabilities of the models on unseen provisions. We provide details on the provisions in Appendix D.

Conclusion

1. The best performing model is not (*yet*) accurate enough
    a. But it can be really helpful in a pre-generation/post-editing scheme
2. Evaluation guidelines were clear, however the legal field is ambiguous and is subject to interpretation
3. Our metric is highly correlated to the human evaluation

Future work

1. Better control the generations of neural networks
   a. Constrained decoding
   b. Weighted beam search
2. Provide more insights during the generation on;
   a. Omissions
   b. Hallucinations
3. See how can our work be integrated with a human in the loop

Thanks!

Nicolas

Eve

Luc

Pierre-Luc

Institut intelligence et données

OBSERVATOIRE INTERNATIONAL SUR LES IMPACTS SOCIÉTAUX DE L'IA ET DU NUMÉRIQUE

SOQUIJ Intelligence juridique

CRSH SSHRC

NSERC CRSNG