



European
University
Institute

DEPARTMENT
OF LAW

ROBERT
SCHUMAN
CENTRE FOR
ADVANCED
STUDIES



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



UNIWERSYTET JAGIELLOŃSKI
W KRAKOWIE



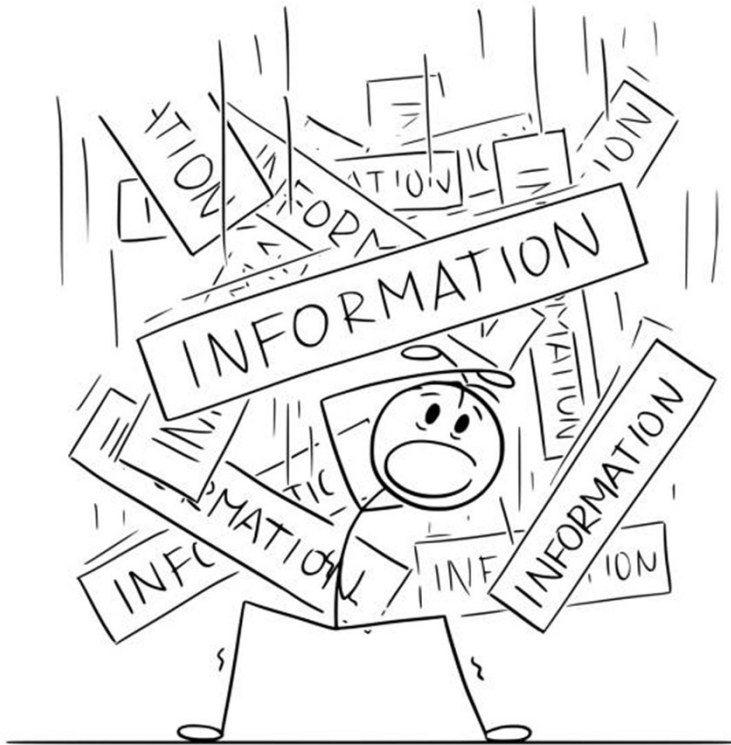
“Make privacy policies longer and appoint LLM readers”:

Reconciling Comprehensiveness and Comprehensibility of Privacy Policies

Przemyslaw Palka, Marco Lippi, Francesca Lagioia, Rūta Liepiņa, Giovanni Sartor

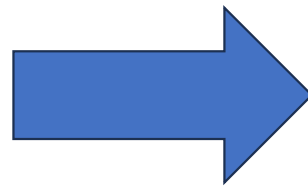
PRIMA final conference – 12 January 2026

The Privacy Policy Landscape



- very long and complex
- do not contain the necessary info
- are full of vague terms (“we collect data about your use of our service”) AND open-ended catalogs (“such as,” “including,” “for example”)
- do not specify what categories of data are shared with whom (“we share your personal data with our marketing partners”)

WHAT IF



LLMs Capabilities

GPT-4 Passes the Bar Exam: What That Means for Artificial Intelligence Tools in the Legal Profession

April 19, 2023 | By Pablo Arredondo, Q&A with Sharon Driscoll and Monica Schreiber

SUBSCRIBE

Can GPT-4 Support Analysis of Textual Data in Tasks Requiring Highly Specialized Domain Expertise?

Jaromir Savelka^{1*}, Kevin D. Ashley², Morgan A. Gray², Hannes Westermann³ and Huihui Xu²

¹Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA

²Intelligent Systems Program, University of Pittsburgh, PA, USA

³Cyberjustice Laboratory, Faculté de droit, Université de Montréal, Montréal, Canada

APPLIED RESEARCH

On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study

DEZHAO SONG¹, SALLY GAO², BAOSHENG HE³, AND FRANK SCHILDER¹

¹Thomson Reuters, Eagan, MN 55123, USA

²Thomson Reuters, New York, NY 10036, USA

³Meta Platforms Inc., Menlo Park, CA 94025, USA

Artificial Intelligence and Law
<https://doi.org/10.1007/s10506-023-09373-8>

ORIGINAL RESEARCH



Multi-language transfer learning for low-resource legal case summarization

Gianluca Moro¹ · Nicola Piscaglia² · Luca Ragazzi¹ · Paolo Italiani¹

Accepted: 5 June 2023

© The Author(s) 2023



LexNLP: Natural language processing and information extraction for legal and regulatory texts

Michael J Bommarito II, Daniel Martin Katz, Eric M Detterman

LexPredict, LLC

LawSum: A weakly supervised approach for Indian Legal Document Summarization

Vedant Parikh^a, Vidit Mathur^c, Parth Mehta^b, Namita Mittal^c, Prasenjit Majumder^b

^aDhirubhai Ambani Institute of Information and Communication Technology

^bParamonic AI

^cMalaviya National Institute of Technology

Legal Requirements for PPs



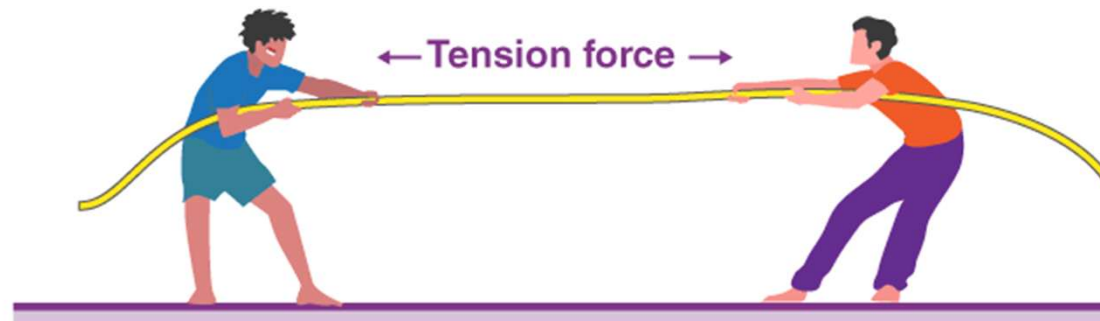
Comprehensiveness vs. Comprehensibility

Comprehensiveness:

A PP should contain all the info relevant for pondering whether to use a service.

Comprehensibility:

A PP should be written in simple language and easy to understand.



The US Approach



- “notice and choice” model:
 - companies who wish to collect and use personal data should make privacy policies available to consumers (“notice”) who, in turn, can decide whether such data practices are acceptable to them (“choice”)
- a matter of contract law and public enforcement by the Federal Trade Commission and state Attorneys General
- Kamala Harris (AG California, 2014):
 - that privacy policies “should provide a comprehensive overview of [...] practices regarding the collection, use, sharing and protection of personally identifiable information” and recommended that drafters of privacy policies “use plain, straightforward language. Avoid technical or legal jargon. Use short sentences”

Comprehensiveness under the GDPR: the EU Approach

Transparency Requirements

Categories of personal data (e.g., email, credit card number)

Purpose of processing for which each category of data is intended (e.g., issuing receipts)

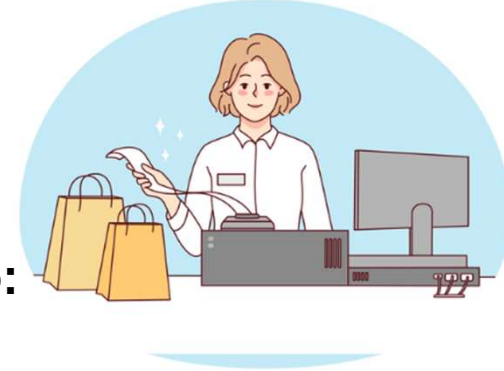
Legal basis for each data processing (e.g., contractual necessity)

Storage period or, when it is not possible, the **criteria** to determine such a period (e.g., till account closure), and with regard to each data and purpose

Recipients or at least the categories of recipients (e.g., a cloud provider), as well as their

- a) Role* (controller or processor)
- b) Purpose* of sharing (e.g., data backup)
- c) Legal basis* of sharing (e.g., legitimate interest)

Assessing Privacy Policies



9 simple questions a consumer should be able to get a clear answer to:

Q1: What data does the company process about me?

Q2: For what purposes does the company use my email address?

Q3: Who does the company share my geolocation with?

Q4: What types of data are processed on the basis of consent, and for what purposes?

Q5: What data does the company share with Facebook?

Q6: Does the company share my data with insurers?

Q7: What categories of data does the company collect about me automatically?

Q8: How can I contact the company if I want to exercise my rights?

Q9: How long does the company keep my delivery address?

Assessing Privacy Policies: Scenarios

- ❑ Scenario 1: Human Evaluation of Existing Privacy Policies
- ❑ Scenario 2: LLMs and Mock Privacy Policy
- ❑ Scenario 3: LLMs and Real Privacy Policies



Scenario 1: Human Evaluation of Existing Privacy Policies

- **Goal:** to what extent legal experts are able to answer the 9 questions from current PPs?
- 5 PPs from food delivery sector (Deliveroo, DoorDash, Glovo, Just Eat, and Wolt)
- Length: 4'379 words (Deliveroo)- 13'765 words (Glovo)
- Each policy independently evaluated by 2 legal experts using the 9 mentioned questions
- **PP Evaluation Criteria:**
 - a) unambiguous (clear) information (i.e., definite answer could be given)
 - b) ambiguous (impossible to know) information (i.e., no definite answer could be provided)
- Each expert were required to add justifications for the evaluations



Scenario 1:Results

Question	Just Eat	DoorDash	Wolt	Glovo	Deliveroo
Q1	X	X	X	X	X
Q2	X	X	X	X	X
Q3	X	X	X	X	X
Q4	X	X	X	X	X
Q5	X	X	X	X	X
Q6	✓	X	✓	✓	✓
Q7	X	X	X	X	X
Q8	✓	✓	✓	✓	✓
Q9	X	X	X	X	X

Table 1: Legal expert evaluation of privacy policies.

Scenario 1: Examples from PPs

- Q9 [Data Retention]: “We will only retain your information for **as long as reasonably necessary** to fulfil the purposes we collected it for, including for the purposes of satisfying any legal, regulatory, tax, accounting or reporting requirements.” [Deliveroo]
- Q7 [Automatically collected data]: “We also **may collect** and store certain information about you and your device(s) automatically when you access or use our platforms and Services, **including, without limitation**, on our websites and engaging with our online chat functionality” [DoorDash]

Scenario 1: Examples from PPs

- Q9 [Data Retention]: “Wolt does not store your ***personal data*** longer than is legally permitted and necessary for the purposes of providing the Wolt Services or the relevant parts thereof.” [Wolt].
- Q2 [Purposes for using email]: “Glovo processes your ***personal data*** for: i) detect and investigate fraud and possible crimes committed against our Platform and all the users, [...].” [Glovo]

The Legal Proposal: Fully Comprehensive Privacy Policies

The law should require corporations to disclose ***fully comprehensive privacy policies*** ideally in a ***standardized form relax the comprehensibility requirement.***

The non ambiguity test: if it is reasonably possible for a reader to ask “**but what do you mean by category X** (e.g., usage or payment information)?”, that category should either **be split** into smaller categories **or exhaustively defined** (e.g., the amount of time you spend on the app and the article you repost; bank account and PayPal email)



The Mock PP: Orderoo INC.

Each paragraph should contain:

- A. data category
- B. source of data
- C. purpose of processing + explanation
- D. legal basis + explanation
- E. storage period
- F. sharing
 - i. the recipient's identity
 - ii. their role (processor and controller)
 - iii. the purpose of sharing
 - iv. legal basis

6880 words long

THIS IS A MOCK POLICY OF "ORDEROO INC."
A COMPANY SIMILAR TO DELIVEROO/ DOORDASH /JUST EAT/ UBER EATS etc.

ORDEROO INC. PRIVACY POLICY

This is a privacy policy of Orderoo Inc., a company located at 1 Name Street, 40121, Bologna, Italy. In this document, we explain what personal data we collect when you are using our service, what source we collect it from, for what purposes we use it, with whom we share it, and based on what legal basis.

We have appointed a Data Protection Officer, who can be contacted at dpo@orderoo.com. You can also contact us by writing to privacy@orderoo.com or at the physical address of our location.

When you use our services, we process the following categories of personal data:

1. Your email address. You provide us with your email address when registering for the service. We use your email address for the following purposes: unique identifier, it serves as a unique identifier allowing you to set up and log in to your account (contractual necessity); account access, to let you reset your password if you forget it (contractual necessity); transaction-related-communication, to send you receipts of your orders (legal obligation: to issue receipts, according to the Receipts Act); distribution of own advertising, to send you advertisements of our own services, new functionalities or new order options (legitimate interest: informing the consumers about the available offers and features, and promoting them); distribution of third-party marketing, to send you advertisements of vendors selling their products on our site (legitimate interest: to subsidize the price of the service with payments from the vendors we promote); tracking transaction history, we keep it as a part of your order history in case it becomes necessary to reveal it to investigative authorities (legal obligation: Accounting Act and Code of

A sample from the fully
comprehensive mock PP.

Scenario 2: LLMs and Mock Privacy Policy

- **Goal:** to what extent are LLMs able to answer the 9 questions from a fully comprehensive privacy policy?
- **Tested LLMs:** GPT-4, Llama-7B, Llama-13B, Llama-70B, Mistral-7B, and Momo-70B (Issue: context window and number of tokens)
- **Selected LLMs:** GPT-4 and Llama2-7b
- **5 iterations** to assess variability across different runs
- Legal expert evaluation of the LLM answers against the Mock PP

The experimental set up

Prompt: In answering the questions please rely solely on the information included in the text and not your knowledge from other sources; please read the document carefully and mention everything, do not omit any information included in the text; please do not shorten or simplify the answers.

Scenario 2: Results on the Mock PP

How to measure performance?

True positive

The reply contains correct information, based on the text

True negative

The reply states that no answer can be provided based on the text (missing info)

False positive

The reply contains incorrect information, though based on the document

False negative

The reply misses information contained in the document

Scenario 2: Results on the Mock PP

Three classic performance results

Precision: fraction of positive predictions, actually labelled as positive = $\frac{TP}{TP+FP}$

Recall: fraction of positive examples that are correctly detected = $\frac{TP}{TP+FN}$

F1: harmonic mean between precision and recall = $\frac{2PR}{P+R}$

Scenario 2: GPT-4 Results on Mock PP



	Precision	Recall	F1	
Q1	100.0	100.0	100.0	
Q2	100.0	100.0	100.0	
Q3	100.0	100.0	100.0	
Q4	96.7	92.0	94.3	Data processed based on
Q5	94.4	90.0	92.1	consent. Data shared with Facebook
Q6	100.0	100.0	100.0	
Q7	100.0	78.2	87.8	Data collected automatically
Q8	100.0	100.0	100.0	
Q9	100.0	100.0	100.0	

- GPT answered the majority of questions correctly (33 of 45 questions)
- It followed the prompt instructions carefully, i.e., it used only the information available in the text and did not shorten or simplify the answers

Scenario 2: GPT-4 Examples from Mock

➤ Q5 [Data shared with FB]

➤ Orderoo: “Data shared with Facebook: default delivery address, each delivery address, details of your order (what you buy), the price you pay, ratings of orders, geolocation, food interests, content preferences, system language, and country of residence.”

➤ Run 4: “Orderoo shares the following data with Facebook: email address, default delivery address, each delivery address, order details, ratings of your orders, your geolocation, content preferences, food interests, system language, country of residence.” **[FN: GPT missed 'the price you pay' . FP: GPT added 'email address' .]**

Scenario 2: LLama2–7B Results on Mock PP

Llama2-7B provided only one (!) correct answer for the 45 questions (9 questions for 5 runs) => impossible to use standard classification metrics

Error analysis:

- FN (missing information) = 91%
- FP (information present in the policy but irrelevant for the question) = 38%
- HAL (answers not based on the analysed policy) = 62%
- SUM (summary) = 27%
- INF (inferences) = 13%



Scenario 2: Llama-7B Examples from Mock

Q9 [Data retention]

- Orderoo: “until you delete your account, plus three months”
- Run 3: “Orderoo keeps your delivery address for 18 months after the last order.” **[HAL]**
- Run 4: “According to the document, Orderoo keeps your delivery address for a period of 18 months after the last order placed. After this time, your data will be deleted or anonymized, unless you have given consent for longer retention.” **[HAL]**

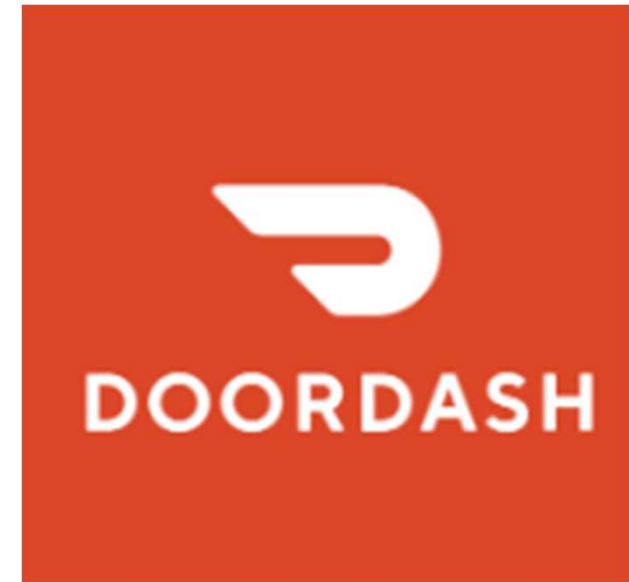
Scenario 2: Llama-7B Examples from Mock

Q8 [Contact Orderoo to exercise rights]

- Orderoo: **“By email: dpo@orderoo.com or privacy@orderoo.com. By mail: 1 Name Street, 40121, Bologna, Italy”**
- Run 1: “[...] you can contact their customer service team at [insert contact information]. Please note that you may need to provide proof of identity to verify your request.” **[MISS]**
- Run 2: “[...] * Email: [info@orderoo.com] (mailto: info@orderoo.com) * Postal address: Orderoo, 123 Main St, Anytown, AB12 3CD.” **[HAL]**

Scenario 3: LLMs and Real Privacy Policies

- **Goal:** to what extent LLMs are able to answer the 9 questions from current privacy policies?
- 5 documents: Deliveroo, DoorDash, Glovo, Just Eat, and Wolt
- Selected LLMs: GPT-4 and Llama2-7B
- Same prompt as for the experiment on the mock policy
- Legal expert evaluation of the LLMs answers against the real PPs



Scenario 3: Results

- GPT-4 provided 9 correct answers, while Llama2-7B only 1 correct answer out of 45 questions in total
- Error analysis:
 - **FN**: GPT-4 47% vs LLama2-7B 64%
 - **FP**: GPT-4 20% vs Llama2-7B 58%
 - **HAL**: GPT-4 0% vs Llama2-7B 38%
 - **SUM**: GPT-4 24% vs LLama2-7B 11%,
 - **INF**: GPT-4 33% vs Llama2-7B 31%

Scenario 3: Examples

Question 3	Who does DoorDash share my geolocation with?
DoorDash PP	iii. Disclosing Personal Information [...] b. Dashers, Merchants or Other Users [...] To help provide you with our delivery services we provide your information, including limited personal information (e.g., first and last names, address, phone number (however, your phone number is masked to protect your privacy)) and order details, to the Dashers who complete your requested deliveries.
GPT-4 Except for Dashers other recipients are missing !!	DoorDash may share your geolocation data with Dashers to facilitate delivery of your orders. They may also use geolocation data to help you select your delivery address, to provide more accurate deliveries, to provide recommendations, and to enhance your user experience. [FN, INF]
Llama2-7B	DoorDash shares users' geolocation data with the following parties: * Third-party delivery partners (e.g., drivers and couriers) * Restaurant partners for delivery (to enable pickup and delivery orders) * Marketing and advertising partners (for targeted promotions and offers) * Service providers (e.g., data analytics and fraud prevention companies). It's worth noting that DoorDash may also share users' personal information with third parties in the event of a merger, acquisition, or sale of assets. [FP, FN, HAL]

Table 4: Comparison of Doordash, GPT-4 and Llama2 answers.

Conclusions

Existing PPs fail to provide meaningful information due to the trade-off between comprehensiveness and comprehensibility

➤ Our experimental setting demonstrated:

- **Scenario 1:** legal experts are unable to answer 7/9 questions due to vague or missing info in the selected PPs
- **Scenario 2:** when given well-structured, fully informative PPs, LLMs (GPT-4 and other LLMs in the future) can provide consumers with meaningful and precise answers.
- **Scenario 3:** when faced with vague and incomplete info, GPT-4 and Llama2-7B produced a high number of FN and FP, SUM, INF, HALL.

In a world of AI-readers it is possible to fully achieve both comprehensiveness and comprehensibility.

The law can and should change!!!



Future Developments

- Expand our experiments:
 - drafting more extensive and diverse fully-informative privacy policies
 - testing additional (reasoning) LLMs
 - linking the LLMs' answers to the relevant portions of the text
 - automating the quality control over such answers.





Francesca Lagioia @eui.eu
@unibo.it

Ruta Liepina @rgsl.edu.lv

Marco Lippi @unifi.it

Przemyslaw Palka @uj.edu.pl

Giovanni Sartor @eui.eu @unibo.it