



Technical challenges

Panel

Mariaceleste Musicco
Giulia Grundler

Purposes

→ **Article 5 (b)** of the Directive requires that: “1. Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;”

→ The ‘**purpose limitation**’ principle under the Article 29 Working Party Opinion 03/2013 on purpose limitation (WP 203) requires that both the purpose for processing of personal data must be known and that the individuals whose data you’re processing must be informed

How to assess the quality of the clause (<purp>)

We assume that each type of required clause informing about the purposes of processing can be classified either as:

- **Sufficiently informative (Level 1):** all the required information - i.e. the purpose of the processing and the personal information processed for that purpose - are present, clear, and well specified (a green clause);
- **Sub-optimally informative (Level 2):** the purpose and required information are present, but information is not fully specified (a yellow clause);
- **Insufficiently informative (Level 3):** the purpose of the processing is vague, and the required information is not well specified (a red clause).

Hierarchical annotation method

Data Specification

Purpose

“We use your your credit card details to enable payment and transaction fulfillment”

Sufficiently informative

Purpose

Data Specification

“Analysing your use and navigation of our Services and Software, as well as your content and metadata about your content through both automated (e.g., machine learning) and manual methods.”

Insufficiently informative

Hierarchical annotation method

The assessment depends on a set of elements and attributes:

➤ **Purpose:**

- Closed Terms: concrete and exhaustive (e.g. “To administer your account”; “to confirm your identity”);
- Open Terms: vague and ambiguous (e.g. “To improve our Service”; “To customize your experience”)

➤ **Specification of Personal Data linked to a specific purpose:**

- Present
- List of examples (non exhaustive):
- Missing

Legal Basis

→ According to art. 13(1)(c) and 14(1)(c), the controller shall provide information related to the legal basis for the processing, where the relevant legal basis must be specified according to art. 6(1) for personal data and to art. 9 for special categories of personal data.

How to assess the quality of the clause (<basis>)

We assume that each type of required clause informing about the purposes of processing can be classified either as:

- **Sufficiently informative:** all the required information - i.e. the legal basis and related purpose(s) of the processing and the personal information processed for that legal basis - are present, clear, and well specified (a green clause);
- **Sub-optimally informative:** the legal basis, related purpose(s) and required information are present, but the information is not fully specified (a yellow clause);
- **Insufficiently informative:** the legal basis, related purpose(s) of the processing is vague, and the required information needs to be better specified (a red clause).

Hierarchical annotation method

Legal Basis

Purpose Specification

“Legitimate Interest relied upon: To provide users with tools to inspire creativity, collaboration,

and enjoyment, and to create opportunities for users to reach new audiences. Information

used: Information You Provide and Automatically Collected Information. [TIKTOK]”

Data Specification

Sufficiently informative

Hierarchical annotation method

The assessment depends on a set of elements and attributes

➤ **Legal Basis:**

- Closed Terms: concrete and exhaustive (e.g. “Consent”)
- Open Terms: vague and ambiguous (e.g. “On legal bases”)

➤ **Specification of Personal Data linked to a specific Legal Basis:**

- Present
- List of examples (non exhaustive):
- Missing

➤ **Specification of Purposes linked to a specific Legal Basis:**

- Present
- List of examples (non exhaustive):
- Missing

Starting from what works well...

As with the <CAT>, BERT models work well in:

- <PURP>/<BASIS> identification: F1 score ~ 0.90
- Sub-element classifications:
 - <Category> Open/Closed: F1 score ~ 0.90
 - <Purpose_Data> Present/Examples/Missing: F1 score ~ 0.85

Challenges

➤ Class imbalance for Level classification

LEVEL	N PURP	N BASIS
1	186	85
2	657	453
3	684	15

Oversampling does not seem to be effective

Oversampling improves results

Impossible to learn for BERT models

Challenges

- Categories with overly broad meanings
- Underrepresented categories

List of closed and open Kind of purposes

1. <u>Gen</u>	18. <u>CustomerSupport</u>
2. <u>GenImprovement</u>	19. <u>GeoService</u>
3. <u>GenResearch</u>	20. <u>SpecFeatures</u>
4. <u>Personalization</u>	21. <u>FreeService</u>
5. <u>Analytics</u>	22. <u>InfoAggregation</u>
6. <u>GenFeatures</u>	23. <u>Contract</u>
7. <u>AccountAccess</u>	24. <u>Consent</u>
8. <u>Functioning</u>	25. <u>LegalObligation</u>
9. <u>SpecResearch</u>	26. <u>VitalInterests</u>
10. <u>ProductRec</u>	27. <u>PublicInterest</u>
11. <u>PriceOffer</u>	28. <u>LegitimateInterest</u>
12. <u>Transaction</u>	29. <u>Tracing</u>
13. <u>Orders</u>	30. <u>AccessComputerInfo</u>
14. <u>Social</u>	31. <u>GenBusiness</u>
15. <u>Verify</u>	32. <u>UserExpectation</u>
16. <u>Ads</u>	33. <u>ServiceCommunication</u>
17. <u>Enforcement</u>	

Challenges

➤ Sparse information

The personal data we receive from Business Partners includes basic personal details, contact details, device details, profile inputs, as well as: **Demographic information**, (e.g. details about age brackets and educational background); **Location data** (e.g. information about postal or ZIP code), and **Purchase information from third party sites** (e.g. information about purchases on other sites).

We will only receive data from our Business Partners where they are legally permitted to share such data, and we will only process that data for the purposes described below. We use the data provided to better understand your preferences, deals that are relevant to you, and how our merchants are performing.

By combining the **data we collect directly from you** with **that received from third parties**, we are able **to provide you with a better, and more personalized, Groupon experience**, as it is in our legitimate interest **to promote offerings you are more likely to buy**.

purp

Category_purp

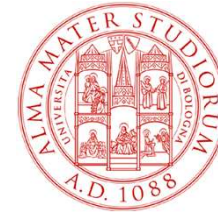
Purpose_data

Challenges

➤ Table format

	How we use your data	Legitimate interests relied on	Personal data used
Safety and security:	<p>To help protect you, us, or others from threats, we log users' activity on LinkedIn to identify and investigate harmful or fraudulent behaviour that violates our User Agreement. Examples of the behaviour include: security threats, hacks, scraping, fake accounts and information, bots, malicious actors, or fraud.</p> <p>To verify certain</p>	<p>To improve the safety, security, authenticity of our Services.</p> <p>To better protect the personal data of our Members and Visitors.</p> <p>To detect, investigate and address threats, bad actors and malicious activity on LinkedIn.</p> <p>So that our Affiliates, including our parent company, Microsoft, can detect, investigate and address threats, possible fraud, bad actors, harmful behaviour and malicious activity against LinkedIn,</p>	<p>Data relating to your use of our Services: IP address, device ID, user agent, location data, browser type and version, operating system and platform, and other online identifiers collected from Cookies and other similar technologies</p> <p>Your account information: your email address, login and two-</p>

- Even if we standardized the format, it is still different from "regular text"
- Removing these sections improves results



Thank you!